PROCEEDINGS OF SPIE

SPIE—The International Society for Optical Engineering

Optical Pattern Recognition XV

David P. Casasent
Tien-Hsin Chao
Chairs/Editors

15–16 April 2004
Orlando, Florida, USA

SPIE PROCEEDINGS SERIES

Volume 5437

TA
1650
.O65
2004
Stacks

TA1650 .O65
Optical pattern recognition.
2004
Received: 2004-06-28 12:15:09
UH MANOA LIBRARY

PROCEEDINGS OF SPIE
SPIE—The International Society for Optical Engineering

Optical Pattern Recognition XV

David P. Casasent
Tien-Hsin Chao
Chairs/Editors

15–16 April 2004
Orlando, Florida, USA

Sponsored and Published by
SPIE—The International Society for Optical Engineering

Volume 5437

SPIE is an international technical society dedicated to advancing engineering and scientific applications of optical, photonic, imaging, electronic, and optoelectronic technologies.

The papers published in this volume compose the proceedings of the technical conference cited on the cover and title page. Papers were selected by the conference program committee, and were subject to review by the editors and program committee. The inclusion of these papers in this publication does not necessarily constitute endorsement by the editors or SPIE.

Please use the following format to cite material from this book:

Author(s), "Title of Paper," in *Optical Pattern Recognition XV*, edited by David P. Casasent, Tien-Hsin Chao, Proceedings of SPIE Vol. 5437 (SPIE, Bellingham, WA, 2004) page numbers.

ISSN 0277-786X
ISBN 0-8194-5360-9

Published by
SPIE—The International Society for Optical Engineering
P.O. Box 10, Bellingham, Washington 98227-0010 USA
Telephone 1 360/676-3290 (Pacific Time) · Fax 1 360/647-1445
http://www.spie.org

Copyright © 2004, The Society of Photo-Optical Instrumentation Engineers

The papers published in these proceedings reflect the work and thoughts of the authors and are published herein as submitted. The publisher is not responsible for the validity of the information or for any outcomes resulting from reliance thereon.

Copying of material in this book for internal or personal use, or for the internal or personal use of specific clients, beyond the fair use provisions granted by the U.S. Copyright Law is authorized by SPIE subject to payment of copying fees. The Transactional Reporting Service base fee for this volume is $15.00 per article (or portion thereof), which should be paid directly to the Copyright Clearance Center (CCC), 222 Rosewood Drive, Danvers, MA 01923. Payment may also be made electronically through CCC Online at http://www.copyright.com. Other copying for republication, resale, advertising or promotion, or any form of systematic or multiple reproduction of any material in this book is prohibited except with permission in writing from the publisher. The CCC fee code is 0277-786X/04/$15.00.

Printed in the United States of America.

Contents

vii *Conference Committee*

SESSION 1 INVITED SESSION ON PATTERN RECOGNITION I

1 **Face recognition with pose and illumination variations using new SVRDM support vector machine (Invited Paper)** [5437-01]
D. P. Casasent, C. Yuan, Carnegie Mellon Univ. (USA)

13 **Portable 512×512 grayscale optical correlator (Invited Paper)** [5437-02]
T.-H. Chao, J. Hanan, H. Zhou, G. Reyes, Jet Propulsion Lab. (USA)

17 **Optical 3D watermarking and authentication by correlation techniques (Invited Paper)** [5437-03]
S. Kishk, B. Javidi, Univ. of Connecticut (USA)

SESSION 2 INVITED SESSION ON PATTERN RECOGNITION II

27 **Efficient method for extracting object recognition metric from correlation filter output (Invited Paper)** [5437-04]
B. V. K. Vijaya Kumar, J. Thornton, Carnegie Mellon Univ. (USA)

38 **Invariant fringe-adjusted joint transform correlation-based target tracking in FLIR sequences (Invited Paper)** [5437-05]
C.-H. Loo, M. S. Alam, Univ. of South Alabama (USA)

SESSION 3 DISTORTION INVARIANT FILTERS

51 **Performance assessment of Unconstrained Hybrid Optical Neural Network (U-HONN) filter for object recognition tasks in clutter** [5437-07]
I. Kypraios, R. C. D. Young, C. R. Chatwin, Univ. of Sussex (United Kingdom)

63 **Performance evaluation for cluttered infrared image based on fringe-adjusted joint transform correlator** [5437-08]
M. S. Alam, E. H. Horache, S. F. Goh, Univ. of South Alabama (USA)

75 **Modified phase-encoded fringe-adjusted joint transform correlation for multiple target detection** [5437-09]
M. R. Haider, M. N. Islam, Bangladesh Univ. of Engineering and Technology (Bangladesh); M. S. Alam, J. F. Khan, Univ. of South Alabama (USA)

87 **On the development of filter management module for grayscale optical correlator** [5437-10]
H. Zhou, Jet Propulsion Lab. (USA); C. Hughlett, Zlabs (USA); J. C. Hanan, T.-H. Chao, Jet Propulsion Lab. (USA)

95 **Performance evaluation of quadratic correlation filters for target detection and discrimination in infrared imagery** [5437-11]
S. R. F. Sims, U.S. Army Aviation and Missile Command (USA); A. Mahalanobis, Lockheed Martin Corp. (USA)

SESSION 4 NEW OPR SYSTEMS AND COMPONENTS

109 **Performance of multilevel error correction in binary holographic memory** [5437-12]
J. C. Hanan, T.-H. Chao, G. F. Reyes, Jet Propulsion Lab. (USA)

115 **Photoconductive optically driven deformable membrane for spatial light modulator applications utilizing GaAs and InP substrates** [5437-13]
B. Haji-Saeed, R. Kolluru, D. Pyburn, R. Leon, S. K. Sengupta, M. Testorf, W. Goodhue, Univ. of Massachusetts/Lowell (USA); J. Khoury, A. Drehman, C. L. Woods, Air Force Research Lab. (USA); J. Kierstead, Solid State Scientific Corp. (USA)

123 **Optical correlator using four kilohertz analog spatial light modulators** [5437-14]
T. Ewing, S. Serati, K. Bauchert, Boulder Nonlinear Systems, Inc. (USA)

SESSION 5 PATTERN RECOGNITION APPLICATIONS I

134 **Distributed multisensor processing, decision making, and control under constrained resources for remote health and environmental monitoring (Invited Paper)** [5437-15]
A. Talukder, Univ. of Southern California (USA) and Jet Propulsion Lab. (USA); T. Sheikh, L. Chandramouli, Univ. of Southern California (USA)

146 **Characterization of photorefractive deconvolution techniques for one-way image transmission through an aberrating medium** [5437-16]
B. Haji-Saeed, S. K. Sengupta, M. Testorf, D. Megherbi, W. Goodhue, Univ. of Massachusetts/Lowell (USA); J. Khoury, C. L. Woods, Air Force Research Lab. (USA); J. Kierstead, Solid State Scientific Corp. (USA)

153 **Resolution limits in imaging LADAR systems** [5437-17]
J. Khoury, C. L. Woods, J. Lorenzo, Air Force Research Lab. (USA); J. Kierstead, Solid State Scientific Corp. (USA); D. Pyburn, S. K. Sengupta, Univ. of Massachusetts/Lowell (USA)

SESSION 6 PATTERN RECOGNITION APPLICATIONS II

163 **Hyperspectral imaging using electro-optic Fourier transform spectrometer** [5437-19]
T.-H. Chao, H. Zhou, Jet Propulsion Lab. (USA); X. Xia, S. Serati, Boulder Nonlinear Systems, Inc. (USA)

171 **Feature selection from high-dimensional hyperspectral and polarimetric data for target detection** [5437-20]
X.-W. Chen, Univ. of Kansas (USA); D. P. Casasent, Carnegie Mellon Univ. (USA)

179 **Fast noninvasive eye-tracking and eye-gaze determination for biomedical and remote monitoring applications** [5437-22]
A. Talukder, J.-M. Morookian, S. Monacos, R. Lam, C. Lebaw, A. Bond, Jet Propulsion Lab. (USA)

191 **Image segmentation using variable threshold functions** [5437-23]
H. Hassan, Lawrence Technological Univ. (USA)

SESSION 7 DISPLAYS, DETECTION, AND TRACKING

203 **Dual LCD-based flat panel display** [5437-24]
J. C. Kirsch, B. K. Jones, U.S. Army Research, Development and Engineering Command (USA)

214 **Improved target detection using polarization-enhanced fringe-adjusted joint-transform correlation** [5437-25]
A. M. El-Saba, M. S. Alam, E. H. Horache, S. Regula, Univ. of South Alabama (USA)

223 **Passive optical methods for helmet tracking in aircraft** [5437-26]
P. J. Vitarius, D. A. Gregory, Univ. of Alabama in Huntsville (USA); D. Morell, FastMetrix, Inc. (USA)

233 **Neural network tracking and extension of positive tracking periods** [5437-27]
J. C. Hanan, T.-H. Chao, Jet Propulsion Lab. (USA); P. Moreels, California Institute of Technology (USA)

238 **Hand motion recovery based on skin color segments and motion parameter estimation** [5437-28]
S. Chang, L. Elkins, Z. Wang, T. Zhou, S. Foo, A. Tchoyprakov, J. J. Fuller, Spatial Integrated Systems, Inc. (USA)

248 **Enhanced projection slice theorem synthetic discriminant functions based on the Karhunen-Loeve transform with application to the protein structure identification in cryo-electron microscopic images** [5437-29]
V. R. Riasati, H. Zhou, California State Polytechnic Univ. (USA)

259 **Optical implementation of image encryption using the zero-order phase-contrast technique** [5437-30]
J. C. Pizolato, Jr., L. G. Neto, Univ. de São Paulo (Brazil)

POSTER SESSION

265 **Classification and rejection of MSTAR data** [5437-31]
L. Ramamoorthy, D. P. Casasent, Carnegie Mellon Univ. (USA)

277 **Face verification and rejection with illumination variations using MINACE filters** [5437-32]
R. Patnaik, D. P. Casasent, Carnegie Mellon Univ. (USA)

288 **Theoretical modeling of the operational mechanism of a photoconductive MEMS spatial light modulator under AC and DC bias** [5437-33]
J. Khoury, C. L. Woods, Air Force Research Lab. (USA); B. Haji-Saeed, S. K. Sengupta, D. B. Megherbi, W. Goodhue, Univ. of Massachusetts/Lowell (USA); J. Kierstead, Solid State Scientific Corp. (USA)

293 **A real-time optical automatic target recognition system** [5437-35]
H. Chen, Southwest China Institute of Electronic Technology (China) and Sichuan Univ. (China); J. Nan, Southwest China Institute of Electronic Technology (China); X. Li, H. Wei, Sichuan Univ. (China)

301 **Incoherent acousto-optic image correlator with the kinoform** [5437-36]
S. N. Starikov, V. G. Rodin, I. V. Solyakin, E. A. Shapkarina, A. P. Chervonkin, Moscow Engineering Physics Institute (Russia)

309 **Design of optoelectronic computer systems for image processing in stationary and dynamic modes** [5437-38]
V. L. Perju, Technical Univ. of Moldova (Moldova); D. P. Casasent, Carnegie Mellon Univ. (USA)

320 *Author Index*

Conference Committee

Symposium Chair

Edward M. Carapezza, DARPA (USA) and Co-Chair DoD/DoJ Joint Program Committee Steering Group for Law Enforcement Related Technologies (USA)

Track Chair

Andrew R. Pirich, Air Force Research Laboratory (USA)

Conference Chairs

David P. Casasent, Carnegie Mellon University (USA)
Tien-Hsin Chao, Jet Propulsion Laboratory (USA)

Program Committee

Mohammad S. Alam, University of South Alabama (USA)
Dennis C. Braunreiter, Raytheon Company (USA)
Don A. Gregory, University of Alabama in Huntsville (USA)
Bahram Javidi, University of Connecticut (USA)
Richard D. Juday, NASA Johnson Space Center (Retired) (USA)
Dennis R. Pape, AlphaLaunch (USA)
Yunlong Sheng, Université Laval (Canada)
Joseph L. Stufflebeam, NewTec (USA)
Ashit Talukder, Jet Propulsion Laboratory (USA)
Bhagavatula V. K. Vijaya Kumar, Carnegie Mellon University (USA)
Rupert C. D. Young, University of Sussex (United Kingdom)

Session Chairs

1 Invited Session on Pattern Recognition I
David P. Casasent, Carnegie Mellon University (USA)

2 Invited Session on Pattern Recognition II
David P. Casasent, Carnegie Mellon University (USA)

3 Distortion Invariant Filters
Rupert C. D. Young, University of Sussex (United Kingdom)

4 New OPR Systems and Components
Tien-Hsin Chao, Jet Propulsion Laboratory (USA)

5 Pattern Recognition Applications I
Ashit Talukder, Jet Propulsion Laboratory (USA)

6 Pattern Recognition Applications II
Tien-Hsin Chao, Jet Propulsion Laboratory (USA)

7 Displays, Detection, and Tracking
Mohammad S. Alam, University of South Alabama (USA)
Don A. Gregory, University of Alabama in Huntsville (USA)

Face Recognition with Pose and Illumination Variations Using new SVRDM Support Vector Machine

David Casasent and Chao Yuan
Dept. ECE, Carnegie Mellon Univ. Pittsburgh, PA 15213

ABSTRACT

Face recognition with both pose and illumination variations is considered. In addition, we consider the ability of the classifier to reject non-member or imposter face inputs; most prior work has not addressed this. A new SVRDM support vector representation and discrimination machine classifier is proposed and initial face recognition-rejection results are presented.
Key words: support vector machines, face recognition

1. INTRODUCTION

We consider face recognition with both pose and illumination variations present and with the need to reject non-member or imposter face inputs. A subset of the CMU-PIE (pose, illumination and expression) database [1] is used (Sect.3). To achieve both good recognition and rejection, a new support vector machine (SVRDM) is used. We recently [2] introduced this new SVRDM system, which we summarize in Sect.2. Our new face recognition results using it are presented in Sect.4.

The FERET database and tests on it have provided much valuable information on face recognition algorithm. However, this database does not include pose or illumination variations. Eigenfaces (view based) [3] have been used to handle pose variations (without illumination variations). Fisherfaces [4] have been noted to be of use for illumination variations, but only frontal pose variations were considered. Many graphics techniques [5-7] have been applied to the PIE and other databases with pose and illumination differences. The light fields method [5] uses 2D images to estimate the 4D light field. The method is computationally expensive on-line; it requires accurate registration and it assumes that a precise pose estimation is available. The illumination cone [6] method uses several different images to produce a 3D surface description of each person. Input faces are matched to the reference face whose illumination cone most closely matches. This method requires much storage. In other work, a morphable 3D face model [7] is produced using a 3D scanner. We assume only 2D sensors. Prior work noted that pose differences are a greater problem than illumination variations [8]. Our results support this.

We chose a support vector machine approach [9,10], since they offer a novel kernel solutions to higher-order classifiers. They also offer good generalization. However, they are not able to reject non-object data, or in the present case to reject imposter faces that are not client members [11-13]. Our support vector representation machine (SVRM in Sect.2.1 is motivated by providing good representation [11-13]. Throughout this work, *we assume that there are no non-object class samples available for training*. This is realistic for face recognition, where it is not realistic to have seen every possible imposter or non-client face. Similarly, in ATR, one cannot expect images of all possible other objects and clutter.

2. SUPPORT VECTOR REPRESENTATION AND DISCRIMINATION MACHINE (SVRDM)

In this section, we describe our new SVRDM classifier. For face recognition and other classification applications, it is vital that the classifier be able to reject non-member inputs. To achieve rejection and to handle true class variations, we first consider one class recognition and our support vector representation machine (SVRM) in Sect.2.1. To extend this concept to include multiple object classes, we present our SVRDM in Sect.2.2. Parameter choices for our SVRDM algorithm are addressed in Sect.2.3.

2.1 SUPPORT VECTOR REPRESENTATION MACHINE (SVRM)

Suppose there are two classes: C_1 is the object class and C_0 is the non-object class (non-member faces for this application). The task of one-class classification is to find the decision region R_1 for C_1 such that if the input $\mathbf{x} \in R_1$, $\mathbf{x}$ is assigned to C_1; otherwise, it is rejected as C_0. Suppose that we are given N training vectors $\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_N\}$ from C_1. *We*

Optical Pattern Recognition XV, edited by David P. Casasent, Tien-Hsin Chao, Proc. of SPIE Vol. 5437 (SPIE, Bellingham, WA, 2004) · 0277-786X/04/$15 · doi: 10.1117/12.548058

assume that no training vectors from C_0 are available in this entire paper; others also make this assumption [11-12]. The training task is to find an evaluation function $f_1(\mathbf{x})$, which gives the confidence of the input $\mathbf{x}$ being in the object class. We define the region $R_1 = \{\mathbf{x}: f_1(\mathbf{x}) \geq T\}$ to contain those object samples $\mathbf{x}$ giving evaluation function values above some threshold T. To achieve a high recognition rate, training vectors should produce high evaluation function values.

We borrow the kernel method used in support vector machines [9,10]. A mapping from the input space to a high-dimensional feature space is defined as $\Phi : R \rightarrow F$, where R is the input space and F is the transformed feature space. The explicit form of Φ and calculation of $\Phi(\mathbf{x})$ are not necessary. Rather, only the inner product $\Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_j)$ need be specified to be some kernel function. To evaluate $\Phi^T\Phi$, we simply evaluate the associated kernel function. We consider only the Gaussian kernel $\exp(-\|\mathbf{x}_i-\mathbf{x}_j\|^2/2\sigma^2)$, since it simplifies volume estimation and has other desirable properties. For a Gaussian kernel, the transformed training and test vectors lie on the unit sphere centered at the origin in F. Since the data are automatically normalized (to be of unit length), the distance between two vectors in F can be represented by their inner product. Thus, as our *evaluation function*, we use the inner product $f_1(\mathbf{x})=\mathbf{h}^T\Phi(\mathbf{x})$, where $\mathbf{h}$ is a vector in F that we compute from the training set. It describes our SVRM and is used to determine the class of test inputs.

The $\mathbf{h}$ solution for our SVRM satisfies

$$\left.\begin{array}{c} \text{Min } \|\mathbf{h}\|^2/2 \\ \mathbf{h}^T\Phi(\mathbf{x}_i) \geq T = 1,\ i = 1,2,...,N \end{array}\right\}. \tag{1}$$

The second condition in (1) insures large evaluation function values for the training set greater than some threshold T (ideally equal to 1). We minimize the norm $\|\mathbf{h}\|$ of $\mathbf{h}$ in the first condition in (1) to reduce the volume of R_1 (to provide rejection of non-objects). We can show that a solution $\mathbf{h}$ with a lower norm provides a smaller class C_1 acceptance volume. In (1), we minimize the square of $\|\mathbf{h}\|$, since such optimization is easily achieved using quadratic programming. In practice, outliers (errors) are expected and we do not expect to satisfy the second constraint in (1) for all of the training set. Thus, slack variables ξ_i are introduced as in SVMs and $\mathbf{h}$ satisfies

$$\left.\begin{array}{c} \text{Min } \left\{ \|\mathbf{h}\|^2/2 + C\sum_{i=1}^{N} \xi_i \right\} \\ \mathbf{h}^T\Phi(\mathbf{x}_i) \geq T = 1-\xi_i,\ i = 1,2,...,N \\ \xi_i \geq 0, \quad i = 1,2,...,N \end{array}\right\}. \tag{2}$$

This allows for classification errors by amounts ξ_i for various training set samples $\mathbf{x}_i$. The factor C in the first condition is the weight of the penalty term for the slack variables. The solution $\mathbf{h}$ to (2) is a linear combination of the support vectors, which are a small portion of the entire training set. To classify an input $\mathbf{x}$, we form the inner product $\mathbf{h}^T\Phi(\mathbf{x})$; if this is $\geq$ some threshold T, we classify $\mathbf{x}$ as a member of the object class.

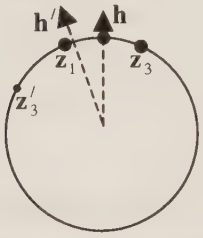


Fig.1 SVRM in the transformed feature space

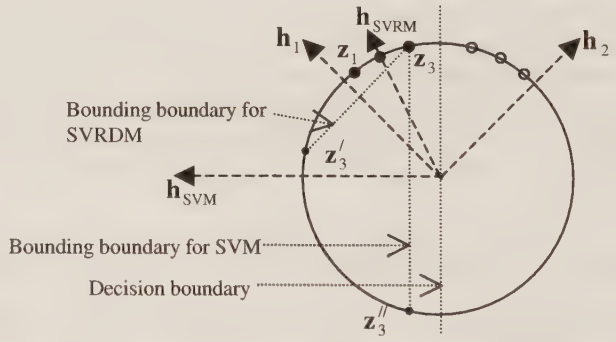


Fig.2. Conceptual SVRM($\mathbf{h}_{SVRM}$), SVRDM($\mathbf{h}_1$) and SVM($\mathbf{h}_{SVM}$) solutions in the normalized transformed space

Fig.1 shows a simple example in which the transformed samples $\mathbf{z}_i=\Phi(\mathbf{x}_i)$ lie on the unit sphere and the training set vectors cover the range from $\mathbf{z}_1$ to $\mathbf{z}_3$. Here, we use a 2D circle to represent the transformed feature space for simplicity. We use Fig.1 to show that an $\mathbf{h}$ solution with a lower norm (or energy) will recognize a smaller range of inputs and is thus expected to produce a lower P_{FA} (better rejection), The vector $\mathbf{h}$ shown crosses the unit circle midway between the end points of the training set data. Its length is such that the inner product $\mathbf{h}^T\mathbf{z}_i$ is ≥ 1 for all $\mathbf{z}_i$. The arc $\mathbf{z}_1\mathbf{z}_3$ in Fig.1 thus

indicates the range of $\mathbf{z}$ that this $\mathbf{h}$ would accept (satisfying $\mathbf{h}^T\mathbf{z} \geq 1$). The solution $\mathbf{h}'$ shown in Fig.1 also satisfies $\mathbf{h}'^T\mathbf{z} \geq 1$; but, the length (norm) of $\mathbf{h}'$ is longer than that of $\mathbf{h}$. This new $\mathbf{h}'$ would result in accepting transformed data over the arc $\mathbf{z}_3'$ to $\mathbf{z}_3$, which is much larger than the extent of the original training set. Thus, use of $\mathbf{h}'$ will lead to more false alarms (FAs) than use of $\mathbf{h}$. Thus, an $\mathbf{h}$ with a lower norm is preferable.

In many circumstances, the training set is not adequate to represent the test set. Thus, in practice, we must use a threshold $T < 1$ in (1) and (2) and must use a decision region that is larger than that occupied by only the training data. However, we do not want this decision region to be too much larger or poor P_{FA} is expected.

2.2 SVRDM ALGORITHM

The SVRM is a one-class classifier that only involves one object class. We now extend the SVRM to the multiple-object-class case. This results in our SVRDM classifier. We consider K object classes with N_k training samples per class; the training vectors for class k are $\{\mathbf{x}_{ki}\}$. We now consider classification and rejection. We define P_C as the classification rate, which is the percentage of the object class samples that are classified in the correct object class. P_R is the rejection rate, defined as the rate of object class samples rejected as the non-object class. P_E is the classification error rate, which is the rate of the object class samples classified in the wrong object classes. Thus, $P_C + P_R + P_E = 1$. P_{FA} is the percentage of the non-object class samples (faces of imposter non-members) mistakenly classified as being in an object class. The objective is to obtain high P_C and a low P_{FA}. Our classifier approach is to obtain K functions $\mathbf{h}_k$; each discriminates one of the K classes (k) from the other $K-1$ classes. For a given test input $\mathbf{x}$, we calculate the VIP (vector inner product) of $\Phi(\mathbf{x})$ with each $\mathbf{h}_k$. If any of these kernel VIPs are $\geq T$, $\mathbf{x}$ is assigned to the class producing the maximum VIP value, otherwise it is rejected. *We assume that there are no non-object class samples in the training set.* For simplicity, we first consider a two-object-class problem. For class 1 samples $\mathbf{x}_{1i}$, we require the evaluation function VIP output $\mathbf{h}_1^T\Phi(\mathbf{x}_{1i}) \geq T$ and $\mathbf{h}_2^T\Phi(\mathbf{x}_{1i}) \leq p$. For class 2 samples $\mathbf{x}_{2j}$, we require $\mathbf{h}_2^T\Phi(\mathbf{x}_{2j}) \geq T$ and $\mathbf{h}_1^T\Phi(\mathbf{x}_{2j}) \leq p$. The parameter p is the maximum evaluation function value we accept for the other object-class samples. The two solution vectors (also referred to as SVRDFs, support vector representation and discrimination functions): $\mathbf{h}_1$ and $\mathbf{h}_2$ must thus satisfy

$$\left.\begin{array}{c} \text{Min } \|\mathbf{h}_1\|^2/2 \\ \mathbf{h}_1^T\Phi(\mathbf{x}_{1i}) \geq 1, \quad i = 1,2,...,N_1 \\ \mathbf{h}_1^T\Phi(\mathbf{x}_{2j}) \leq p, \quad j = 1,2,...,N_2 \end{array}\right\}, \qquad (3)$$

$$\left.\begin{array}{c} \text{Min } \|\mathbf{h}_2\|^2/2 \\ \mathbf{h}_2^T\Phi(\mathbf{x}_{2i}) \geq 1, \quad i = 1,2,...,N_1 \\ \mathbf{h}_2^T\Phi(\mathbf{x}_{1j}) \leq p, \quad j = 1,2,...,N_2 \end{array}\right\}, \qquad (4)$$

Note that the VIP kernel function value for the object class to be discriminated against is specified to be p in this case. We typically have selected p in the range $[-1, 0.6]$. For this face database with all facial parts aligned, we used a much lower p=0.2 value. If we use $p = -1$, then (3) and (4) describe the standard SVM. The classifier in (3) and (4) is our new SVRDM. The difference in the formulation of the SVRM and the SVRDM lies in the third condition in (3) and (4); this condition provides discrimination information between object classes by using $p > -1$ and use of $p < -1$ (the SVM solution is $p = -1$) and rejection of non-objects (imposter not-member faces). In the presence of outliers (training class errors), slack variables ξ_i are of course used in both $\mathbf{h}_1$ and $\mathbf{h}_2$. The final version for $\mathbf{h}_1$ is thus

$$\left.\begin{array}{c} \text{Min}\left\{\|\mathbf{h}_1\|^2/2 + C\left(\sum \xi_{1i} + \sum \xi_{2j}\right)\right\} \\ \mathbf{h}_1^T\Phi(\mathbf{x}_{1i}) \geq 1 - \xi_{1i}, \quad i = 1,2,...,N_1 \\ \mathbf{h}_1^T\Phi(\mathbf{x}_{2j}) \leq p + \xi_{2j}, \quad j = 1,2,...,N_2 \\ \xi_{1i} \geq 0, \xi_{2j} \geq 0, \end{array}\right\} \qquad (5)$$

and $\mathbf{h}_2$ is similar.

To illustrate the difference between different SVM machines, Fig.2 conceptually shows several $\mathbf{h}$ vectors for class 1 and 2 training data noted in transformed space by filled-in and open circles. We show the conceptual solutions $\mathbf{h}_{SVRM}$ (for the SVRM), $\mathbf{h}_1$ (for the SVRDM) and $\mathbf{h}_{SVM}$ (for the SVM). For simplicity, only 3 training data points are shown for each class; they denote the extremes of the training set and the average training set vector for each class. The SVRM solution for class 1 is the vector $\mathbf{h}_{SVRM}$ that passes through the average class 1 vector. The SVM solution $\mathbf{h}_{SVM}$ uses $p = -1$. It is normal to the decision boundary, which is a vertical line through the center of the circle. For our new SVRDM with $p = 0.6$, the vector $\mathbf{h}_1$ shown satisfies (3). The $\mathbf{h}_2$ solution vector for class 2 in (4) is similar to $\mathbf{h}_1$ but it lies in the first quadrant, symmetric about the vertical for this example. The lengths of the three vector solutions shown are proportional to their norms. We note that the energy of $\mathbf{h}_{SVM}$ is more than that of $\mathbf{h}_1$ and $\mathbf{h}_2$ which in turn are more than that of $\mathbf{h}_{SVRM}$. These results are all expected. The range of transformed inputs that will be accepted into class 1 for the SVRDM are described by the arc $\mathbf{z}_3'$–$\mathbf{z}_3$ in Fig.2. This range is much less than the $\mathbf{z}_3''$ to $\mathbf{z}_3$ range for the SVM. Thus, we

expect a lower P_{FA} for our SVRDM vs. the SVM. If Fig.2 were a hyper-sphere, the bounding boundary for the SVM would be a hyper-circle on the surface of the hyper-sphere.

For an N-class problem, we form N pairs of SVRDFs. Each pair is intended to recognize one class vs. the rest, thus class 2 is the rest of the classes in the above 2 class example. For a given test input, if any evaluation function VIP gives a value above the threshold T, the input is classified to the associated class. If no evaluation function VIP gives an output above T, the test input is rejected as a non-object.

2.3 SVRDM PARAMETER SELECTION

The SVRDM has several parameters to be chosen. In our work, we fixed C in (2) and (5) at 20, since we typically had few non-zero slack variables. In other cases, other C choices are preferable. For p, we have typically used p=0.6 for tests on synthetic data.

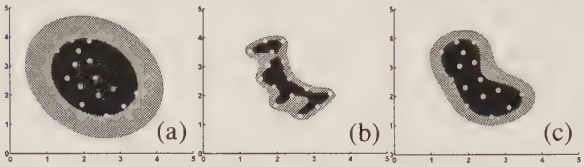


Fig.3 Bounding boundary and bounded region of the SVRM (a) σ=2.0 (b) σ=0.3 (c) σ=0.8

We now discuss selection of σ. We found performance to be significantly affected by the choice of σ in the Gaussian kernel function. For a one class case with 15 samples, Fig.3 shows how the region of space (in which samples will be called members of the object class) changes with σ. The dark region corresponds to the set of $\mathbf{x}$ value for which the VIP (vector inner product) $\mathbf{h}^T\Phi(\mathbf{x}) \geq 1$. We refer to this as the *bounded region* (BR). The outer extent of this region is the *bounding boundary* (BB). As seen, as we reduce σ, the bounded region shrinks and data in a smaller range of the input space is recognized as the object class. The data points on the boundary of the dark region are support vectors. For σ = 0.8, few (only 6 – 8) of the 15 samples are support vectors (Fig.3c). At the lowest σ = 0.3 choice (Fig.3b), all training samples are support vectors on the dark region boundary, and in this case we expect SVM concepts to be lost and test set performance to be poor. For larger σ choices (Fig.3a), the dark region is noticeably larger than for intermediate σ choices (Fig.3c). We thus expect that the σ = 2.0 choice in Fig.3a will yield more false alarms and poorer rejection. For the cases we consider, we expect to have to accept data in a region of space larger than that occupied by the training set data to achieve acceptable test set performance. The boundary of the lighter gray area in Fig.3 shows how much of the input space satisfies VIP output values $\mathbf{h}^T\Phi(\mathbf{x}) \geq$ a lower threshold 0.8. As seen, larger σ values produce a proportionally larger area in the amount of space allocated to the object class. Thus, we expect the intermediate σ value (Fig.3c) to produce better acceptance and rejection results than a high value of σ (Fig.3a).

Prior work has used cross-validation to select σ [7]. We cannot use this since there are no non-object samples in the training set. The goal is to select σ to produce a tight bounding boundary around the N training samples. For a simple data distribution such as a single cluster as in Fig.3, we can easily choose σ to simply be d_a, the average distance of each training sample to the object class center (the mean of all the training samples) as in Fig.2c. However, much real data have distributions that have a complex shape and multiple clusters and in such cases selecting σ is difficult. The general idea in our algorithm is to first obtain an approximate bounding boundary for the object class. To do this, we compute N local SVRMs, since we can easily choose σ for each of these. Each local SVRM has its own local bounding boundary. We analyze these to obtain a set of sample vectors {v} that lie on an approximate boundary for all of the data. In step 2, we compute an SVRM for all of the training data using different σ values. Each will produce a different bounding boundary and a different set of training vectors {w} will lie on each boundary. The σ choice for which the set of training vectors {w} and {v} best matches is the σ value we use. We recently [2] detailed this algorithm.

Fig.4 shows a two-class example in the 2D input space rather than in the unit-sphere normalized transformed space in Fig.1. There are 15 training vectors for each class (solid and open circles). Figs.4a to d show the results for the 3 different machines as p is varied. The bounded region for which the VIP $\geq$ 1 for each case is shown by the dark gray region. The decision boundary for which the two VIPs are equal ($\mathbf{h}_1{}^T\Phi(\mathbf{x}) = \mathbf{h}_2{}^T\Phi(\mathbf{x})$) is shown by the dotted line. As seen in Fig.4a, some class 1 samples lie on the wrong side of the decision boundary and represent errors for the SVRM; this is expected since the SVRM is not designed to provide two-class discrimination. The other three support vector classifiers in Figs.4b to 4d produce good and similar decision boundaries. We thus expect the SVRDM and the SVM to have similar discrimination capabilities. The SVRM produces the smallest dark gray bounded region in which objects are expected to be classified with high confidence. This region for the SVRDM fits the data well and its extent is much

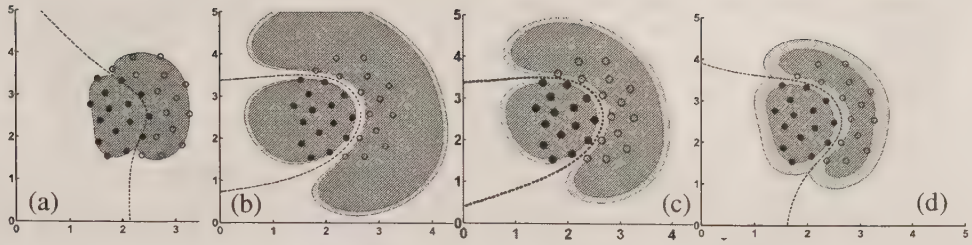


Fig.4 Bounding boundary, bounded region and decision boundary of the (a) SVRM, (b) SVM, (c) SVRDM (p = 0.2) and (d) SVRDM (p=0.6)

less than that produced by the SVM (Fig.4b). As we vary p, we control the characteristics of the classifier (its acceptance and rejection range). With a larger dark gray bounded region of acceptance, we expect *the SVM to yield more false alarms than the SVRDM* and that *both classifiers will produce similar discrimination performance (determined by the decision boundary)*. The p = 0.2 to 0.6 choices result in a smaller dark gray bounded region than does p=-1 for the SVM.. The distribution of non-object to be rejected will determine the best p choice. If non-object lie near the decision boundary between classes, then a lower p value is needed. We expect to reduce the threshold T below one for our applications, as noted earlier; this is expected since very likely that test data will be distributed near the training set data, but will lie outside the dark gray bounded region. The light gray regions in Fig.4 indicate the outer bounds of the regions of the 2D input space for which the kernel VIPs are ≥ 0.8. As we gradually reduce T, the decision region will become larger. Since the bounded region for the SVRDM better fits the distribution of the training data, we expect better P_{FA} vs. P_C results.

3. DATABASE AND SCORING

3.1 PIE FACE DATABASE

The subset of the PIE database [1] that we use consists of 21 illumination variations for each of 13 pose variations for each of 65 subjects. There are thus a total of 17,745 face or 273 images per person. Fig.5 shows the 13 pose views and the pose numbers we use for one subject. The arrangement follows the camera view; poses 1 to 9 vary in aspect angle from left to right pose, poses 10 and 11 are frontal at different elevation angles and poses 12 and 13 are at different elevation angle in corners of the room as in surveillance camera views. The 21 different illuminations are shown in Fig.6. The images we use are 60×48 pixels, with resolution reduced by averaging adjacent pixels in the original 640×486 pixel image by linear interpolation.

3.2 TRAINING AND TEST SETS

Our training set consists of nine pose variations (aspect view poses 1–9) and four illuminations (2, 15, 18, 20) for each. The remaining four poses (elevation angle differences) and 17 illuminations for each face are the typical test set. There are thus typically 9×4=36 training images per person and 4×17=68 test images per person. We use 25 of the persons as a non-object (imposter, non-client) database of faces to be rejected and include all 21×13=273 images of each of these persons (6825 images) in our non-object test reject classes. We varied the number of persons in the object class from 10 to 40.

3.3 REGISTRATION PREPROCESSING

All pose and illumination variation images for each person at a given pose were taken over a very short 0.7 second time interval, during which there is basically no head subject movement. A number of subjects have their heads tilted (in all images), e.g. Fig.7a. If such tilted images of this and other subjects were used in training and recognition, biased results with higher P_C scores would result with recognition based on the subject's tilt angle in addition to his facial features. Thus, to provide unbiased testing, we use a registered (aligned) training set of images of all faces. This registration was also used in previous work on this database [5, 6, 8]. We refer to this as face registration. For each facial pose, we locate the positions of several face landmarks (the eyes and mouth center and for the four feature case, the tip of the nose). We note that we found that the location accuracy of the last nose tip feature is poor. These spatial features are presently located manually, although methods to locate eyes [14] and other [15] facial features have been proposed. For each pose, we normalize all training set images to have the same eye spacing, and eye, nose and mouth positions.

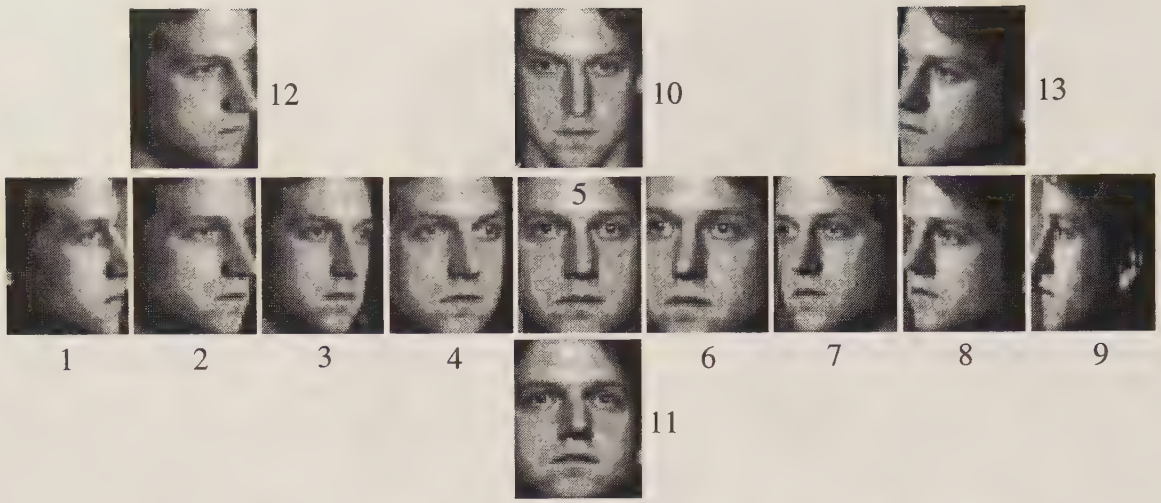


Fig.5 Sample registered images showing pose variations.

Fig.6 Sample registered images showing 21 illumination conditions

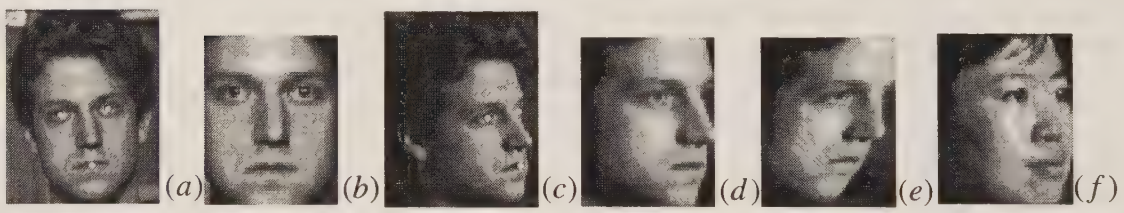


Fig.7 Face registration; original (a) and transformed (b) example. Profile example: original (c), transformed using eyes and mouth landmarks (d); and distorted image using eyes and nose tip landmarks (e). (f) the original image at pose 2 for the person used to select the normalized facial landmark locations.

For each pose, the eye separation is different for a different pose, but the vertical location of the eyes, mouth, etc are the same for all cases. This removes the tilt bias. However, it makes all images more similar and thus makes discrimination and rejection more difficult.

We now detail how to select the normalized locations of the facial landmarks for each pose. Fig.7f shows the original image (before registration) at pose 2 for the person used to select the facial landmark normalized locations for each pose. The subject chosen for each pose was one whose head was not noticeably titled in the head-on frontal-view image (pose 5). Since pose 2 and pose 8 are symmetric, we desire that the normalized locations of the landmarks in the pose 2 and

pose 8 images are also symmetric. However, this symmetry is typically not true in original images. The 2 eyes have different vertical distances in pose 2 and pose 8. Similarly the distance of the eyes, nose and mouth from the edge of the image is different in the true images (the head is not usually centered and is slightly titled). To shift the reference locations of the facial landmarks, we shifted the locations of the eyes vertically and horizontally (so that symmetric eyes, e.g the left eye in pose 2 and the right eye in pose 8 lie on the same vertical row, and so that they are the same distance from the edge of the face and so that they are the same distance apart in the symmetric images). We shifted the locations of the nose and mouth horizontally and vertically similarly. The shifts of all landmark locations are typically only 1-2 pixels horizontally and vertically. This was done for each symmetric pair of poses and for the frontal pose 5 for our 9 standard poses 1-9.

For the entire training set, we performed this image registration separately for each pose. This was done using affine transforms. These affine transforms are preferable as they handle the more complex realistic cases beyond simple object rotations and scales [16]. We locate the eyes and the center of the mouth (see white dots in Fig.7a). Our affine transforms are expressed by

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}, \tag{6}$$

where $[x', y']^T$ is the transformed point and $[x, y]^T$ is the point before the transform. a_{ij} (i=1,2 and j=1,2,3) are the components of the affine transform matrix and are estimated using training set data to minimize MSE (mean squared error).

We retain the face from the forehead to the chin (see Fig.7b). Fig.7b shows the registered version of Fig.7a with the tilt removed. Fig.7c shows a half-profile image and its registered version using the eyes and mouth center as transform feature points. If the eyes and tip of the nose are used and the position of the nose tip is fixed in the registered image, then poorer distorted registered images such as Fig.7e can result. This is attributed to the large variations present in the length and size of noses and in the accuracy with which the nose tip can be located. Fixing the location of the nose tip in the registered image thus often distorts the lower part of the face.

4. FACE RECOGNITION TEST RESULTS

4.1 ONE-STEP INTERPOLATION AND EXTRAPOLATION FACE CLASSIFICATION

In our prior work [2], we used ten faces as our object classes and ten others as the non-object class to be rejected, eight approximately evenly-spaced illuminations were included in the training set and the remaining 21−8=13 illuminations were included in the test set. Pose variations are much more of a challenge than illumination variations [8]. In our *interpolation* experiment, five alternate poses among the central nine (poses 1, 3, 5, 7, 9) were used as the training set and the other eight poses as the test set. There are thus 8×5=40 training images per face and 233 test images per face. For the non-object class test faces, there were 21×13=273 images per faces. For the SVRDM, we set σ=2000 using our new algorithm (Sect.2.3). For the SVM, we set σ=1800. We compared our results to those using Eigenface [3] and Fisherface [4] methods. For the Eigenface method, we used a total of 80 eigenvectors; and for the Fisherface method, we used K-1 Fisher vectors (for K=10 classes). Each Fisher vector considers one true face class vs. all of the other classes. We use the covariance matrix (we subtract the mean of all images), thus the rank of the numerator in Fisher is K-1, not K. We project the test input onto these K-1=9 vectors and classify the test input using a nearest-neighbor classifier. A nearest-neighbor classifier was used for both the Eigenface and Fisherface features. Fig.8a shows ROC (*receiver operating characteristic*) results. P_C is the fraction of the test set object class samples correctly classified. P_{FA} is the fraction of the non-object samples not rejected (errors). In the extrapolation tests, poses 1-5 from only one side were used in the training set, and the illumination training set was unchanged. For the SVRDM, σ=2000 was used and σ=1750 for the SVM. Fig.8b shows ROC results. The EER for the SVM and the SVRDM are 29% and 37%, respectively (equal error rate (EER) is the ROC point in the top left where P_E=1−P_C=P_{FA}). The threshold T was varied to produce the Fig.8 ROC results.

In both ROC cases in Fig.8, the SVRDM performs best, followed by the SVM. The SVM required 151 support vectors per person, while the SVRDM needed only 95. Thus, the on-line computation requirements for the SVRDM are much less than for the SVM. In Table 1, we list the highest P_C performance P_{HC}, ignoring rejection performance. This is not our main concern here, but it is interesting to note that P_{HC} for the SVM and SVRDM are comparable and better than P_{HC} for both the Eigenface and Fisherface classifiers. We also note that performance is better for interpolation (when

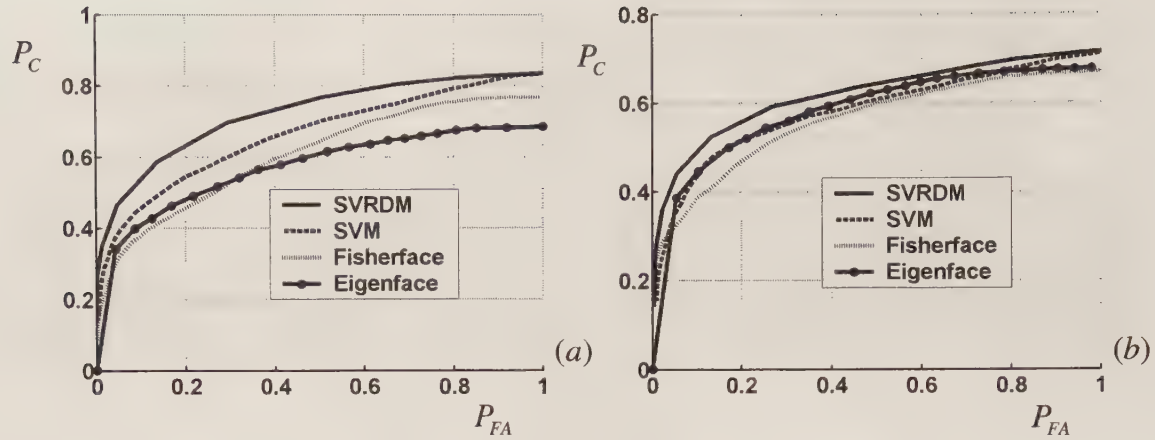


Fig.8 Experiments on PIE database. P_{FA} vs. P_C curves for (a) the interpolation test and (b) the extrapolation test for the one-step method

only intermediate poses are missing) than for extrapolation. We also note that the Fisher classifier is better than the Eigen one (as claimed in [4]) only in the interpolation case.

P_C	Eigen	Fisher	SVM	SVRDM
interpolation	0.685	0.767	0.836	0.835
extrapolation	0.676	0.669	0.708	0.713

Table 1. Highest P_C performance for the one-step method

To obtain good performance on the 40 class face recognition case with 25 non-object faces to be rejected, we only consider interpolation and we increased the number of pose views in the training set to include poses 1-9 (only poses 10-13 with a different elevation angle are now in the test set). Since illumination difference are more easily handled, we train on only illuminations 2, 15, 18 and 20, with the other 17 illumination being the test set. There are thus 9×4=36 training set images for each of the 40 faces and 273−36=237 test set images for each face. For the 25 faces to be rejected, there are 273 versions of each. We refer to these as our *one-step training and test sets*. The SVRDM performed best with EER=0.216 and P_{HC}=87.2%. The SVM performed worse and the other classifiers were even poorer. Most errors were due to pose variations.

4.2 TWO-STEP FACE CLASSIFICATION

To improve results, we consider a two-step algorithm. In step 1, we estimate the pose of the input face; and in step 2, we estimate the class of the input face. In step 1, there are nine classes (each is one of the training set poses 1-9). The same four illuminations (2,15,18,20) were used in the training set. We consider only the SVRDM for step 1. We produced nine SVRDFs; each was one class (one pose) vs. the other eight and each used all nine (poses) × four (illuminations)×K (object classes) = 36K training set images (we varied K and kept 25 faces to be rejected). In step 1, pose estimation, the inner products of the test input and the set of support vectors were separately calculated for each SVRDF and a set of evaluation function values were obtained. The largest evaluation function value determines the class (pose) of the input. *In step 1, there is no threshold T for rejection and the best pose estimate is accepted.* The test set consists of poses 10-13 at all illuminations and poses 1-9 at 17 illuminations for K faces, plus 273 versions of the 25 faces to be rejected. Pose estimation (step 1) actually contains 2 steps of classifiers. This was necessary because the input test data could not be registered without having a pose estimate and an accurate pose estimate required registered test set data. In stage 1 of pose estimation, we used the locations of four facial landmarks (2 eyes, nose and mouth) as the features for our stage 1 SVRDM. These facial landmarks can be automatically located [14, 15]. The SVRDF evaluation functions with the largest 3 values were retained (i.e. we retained the 3 top pose estimates, since the top pose estimate was often not correct). All data in stage 1 were not registered. The input test face was then transformed and registered to each of the top 3 pose estimates from stage 1 and these 3 registered faces were then fed to the stage 2 pose estimator. The stage 2 pose estimator in step 1 has a set of 9 SVRDF functions, one for each pose. It uses registered training set images and iconic features. The largest evaluation function output determines which of the top 3 pose estimates from stage 1 is the final pose estimate used. In both step 1 pose estimation classifiers, no rejection considered, as we desire an

estimate of the pose of the input, rejection is addressed in our step 2 classifier. This transformed and registered input test face and its pose estimate is fed to the step 2 classifier.

In step 2, there are 9 classifiers, one for each pose and each is trained only on objects at the given pose. Each has K SVRDFs, one per class; each is intended to recognize one class and discriminate it vs. all K-1 other face classes. Classes are now a person's face at a given pose. For a given pose, the set of K person (face) functions are evaluated, the largest $\geq T$ determines the face class. If no output is $\geq T$, that test input is rejected as a non-object (in this case an imposter, a person not to be admitted).

We expect better performance from this two-step classifier, since each classifier is now simpler. For the step 2 classifier, we compare 4 different classifiers. Recall that there are two stages in step 1 as in Sect.3.3. For the stage-1 SVRDM, we use σ=0.095. This σ value is much lower, since this stage 1 classifier uses only the normalized coordinates of four facial features and not iconic features. The average number of support vectors per SVRDF in this first stage 1 is 30. The training set gave P_c=99.3%. The average number of slack variables per SVRDF in this stage 1 is 3. For the stage-2 SVRDM classifier in step 1, we use σ=2455. This σ value is much higher than that used in stage-1 classifier, since we now consider quite different iconic features. The average number of support vectors per SVRDF in stage-2 classifier in step 1 is 142. The average number of nonzero slack variables for each SVRDF function was zero in the second-stage classifier in step 1 (thus, the training set gave P_c=100%). This occurs because in the high-dimensional iconic feature space, it is easier to separate different pose classes, probably because images are registered. We use p=0.6 for both classifiers in step 1, since this is a classification not a rejection problem. The choice of p is not critical.

For the SVRDM and the SVM in step 2, we used a different σ for each pose, because the σ we obtained for each SVRDM (for each pose) was quite different (from 1750 to 2600). Thus we did not average them as we did in the step-1 cases. We used p=0.2 in our step-2 SVRDM. We generally pick p=0.6 for an SVRDM. However, the non-objects for our face recognition problem are also faces; these non-object faces are not distributed uniformly in the entire iconic feature space, but rather are only in a small face subspace. Thus, p=0.6 is not the best choice for this particular face recognition case. To better determine the choice of p, we used a validation set which contains 5 faces as object faces and another 5 faces as non-object faces from the PIE database. We varied p from -1 to 1 in increments of 0.2 and found p=0.2 produced the best EER scores for the validation set. This p=0.2 choice is used for both the SVRDM and the SVM in step 2.

For step 2, we also consider an Eigenface (view-based) classifier. We fixed the number of eigenvectors per pose class to be 20, since they contain > 90% of the total energy of the training set images at a given pose (when K=40 classes). We also consider the Fisherface classifier. For the Eigenface and Fisherface features, we use a nearest neighbor classifier.

We increased the number of classes (faces) to be recognized from K=10 to 40 with the number of non-object classes (false faces) to be rejected fixed at 25. The number of support vectors increased as K increased and the SVM always required more support vectors than the SVRDM (26 vs. 18 at K=10 and 46 vs. 25 at K=40). This is the average number of support vectors per pose and per face in step 2. Table 2 shows EER results for the four different K choice and for the four different classifiers. As the number of classes (K) increases, performance of the two support vector machine does not vary noticeably. Thus, they handle larger problems well. However, this is not the case for the other classifiers. They degrade as K increases. The SVRDM again performs the best. The Eigenface classifier performs especially poor, since Eigenfeatures are not intended for discrimination. The EER value noted in Table 2 is one point on the ROC curves. Fig.9 shows the ROC curves for the K=40 class case. If we ignore rejection, which much work does, then the P_{HC} performance at K=40 was as follows: SVM (93.1%), SVRDM (92.6%), Fisherface (88.2%), Eigenface (61.5%).

EER	K=10	K=20	K=30	K=40
SVRDM	15.5%	14.4%	14.3%	15.0%
SVM	22.6%	18.3%	18.4%	18.9%
Fisherface	15.2%	19.1%	22.1%	23.4%
Eigenface	38.7%	46.7%	50.9%	52.0%

Table 2 Equal error rates of different classifiers for different number of object faces

This P_{HC} performance for the SVM and the SVRDM is better than that in Sect 4.1 by ~10%. The EER for the SVRDM is better by ~6%. Previous graphics methods typically used one image per face in training and tested versus all other face images (under different poses and illuminations). 34 faces were used to construct the Eigen light fields [5]; for the remaining 34 faces, one image per face was included in the training set (thus, there were 34 training set images) and the rest of the images were used in testing. A very poor P_C=36.0% was obtained [5] in this prior work. A different subset of the PIE database (with the room lights on and with only three pose variations considered) was used in the 3D

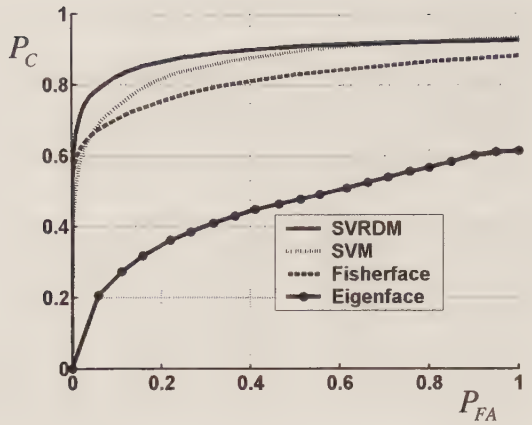


Fig.9 ROC curves for different classifiers using 40 object faces and 25 non-object faces

Morphable model method [7]. P_C=81.1% was obtained for the Morphable model method. Our result is better than these two graphics methods and we considered more pose and illumination variations. In addition, none of these graphics method considered face rejection. The performance of the SVRDM is the best, but it is not excellent. This is somewhat expected, since the pose of the test inputs are not one of the training set poses; they vary in elevation. Thus, we might expect poor pose estimation accuracy for the test set. Similarly, there are 17 illumination variations present in testing and only four present in training and this is expected to affect results .As expected, if the training set doesn't describe the test set, performance will suffer.

To address such training set issues, we consider only the K=40 class case and its SVRDFs. Initially, it gave an EER=15.0% and P_{HC}=92.6%. If we increase the number of illumination differences in the training set to 8 (illuminations 2, 4, 6, 8, 10, 12, 14 and 16), we find EER=11.7% and P_{HC}=95.8%. These are significant improvements. If we include all 13 poses in the training set (but only the original four illumination differences), we obtain very significant improvement to EER=4.2% and P_{HC}=98.9%. Thus, *pose variations are much more significant than illumination variations*. There was only one slack variable in the original test (K=40). There are zero and one slack variable in the last two cases, respectively. Thus, for example, a face image at poses 2 and 12 may seem very similar, but to a classifier they are quite different. In the original 2-step data, P_C=81.1% for input poses 10-13 and P_C=96.7% (illumination) for poses 1-9.

4.3 FACE VERIFICATION

In verification, the person states his identity and enters his facial image. If the match with the stated person is above some threshold T, the person is accepted; this contributes to the true acceptance rate (TAR) or P_C which is the fraction of all true test inputs N that were accepted. We wish to evaluate the usefulness of different classifiers for verification. In this case, we vary the output threshold T for the evaluation functions for the different classes for each classifier. In tests of non-object or imposter face inputs to be rejected, FAR *or* P_{FA} are (as before) the number of faces not rejected.

We use K=40 object faces and 25 non-object faces to be rejected with the same 9 poses and 4 illuminations used in the training set and with 273−36=237=N_1 test set inputs used per object class face and N_2=273 test inputs used for non-object test set inputs. All test set inputs are as before. Fig.10 shows ROC data P_C (or TAR) vs. P_{FA} (or FAR). At P_{FA}=0.1, the TAR rates are 98.5% (SVRDM), 96.0% (SVM), 90.7% (Fisherface), and 65.3% (Eigenface). The EER values are 4.0% (SVRDM), 6.0% (SVM), 10.0% (Fisherface) and 20.9% (Eigenface). Thus, the SVRDM performs best; the Eigenface processor performs worse.

SUMMARY

We presented test results for face recognition using the CMU PIE face database; both pose and illumination variations were considered. The classifiers investigated include the SVM, the SVRDM and two other classic methods: the Eigenface and Fisherface classifiers. We considered two strategies for this face rejection-classification problem. One approach used a view-based two-step strategy, in which the pose of a test input is first estimated and this is followed by an identity classification processor assuming the estimated pose. Our experimental results showed that the SVRDM performs best among all classifiers using the two-step strategy and that the SVRDM was less sensitive to the size of the classification problem than were other classifiers. Our results also note that it is necessary to include more pose and

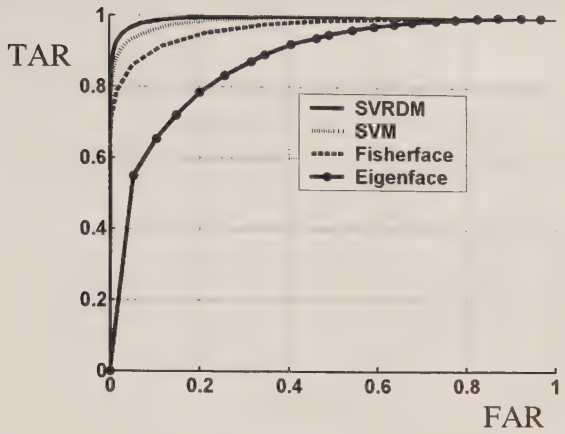


Fig.10 Verification ROC curves for different classifiers.

illumination variations in the training set, so that the training set is more representative of the test set. Furthermore, pose variations are found to be a more challenging problem than illumination variations. Our SVRDM was also shown to be applicable to the verification problem and it was shown to handle the unseen imposter case better than other classifiers do.

REFERENCES

[1] T. Sim, S. Baker and M. Bsat, "The CMU Pose, Illumination, and Expression (PIE) Database", IEEE. Conf. on Automatic Face and Gesture Recog., May, 2002, pp.46-51.
[2] C. Yuan and D. Casassent, "Support Vector Machines for Class Representation and Discrimination", Int'l Joint Conf. on Neural Networks, Portland, July, 2003.
[3] A. Pentland, B. Moghaddam and T. Starner, "View-Based and Modular Eigenspaces for Face Recognition", CVPR, 1994.
[4] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection", IEEE Trans. on PAMI, Vol.19, No.7, July, 1997, pp.711-720.
[5] R. Gross, I. Matthews and S. Baker, "Appearance-Based Face Recognition and Light-Fields", CMU-RI-TR-02-20, Aug. 2002.
[6] A. S. Georghiades, P. N. Belhumeur and D. J. Kriegman, "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose", IEEE Trans. on PAMI, Vol.23, No.6, June, 2001, pp.643-660.
[7] V. Blanz, S. Romdhani and T. Vetter, "Face identification across different poses and illuminations with a 3D morphable model", IEEE Conf. on Automatic Face and Gesture Recog. 2002, pp.192-197.
[8] R. Gross, J. Shi and J. F. Cohn, "Quo vadis Face Recognition", Third Workshop on Empirical Evaluation Methods in Computer Vision, Dec. 2001.
[9] C.Cortes and V. Vapnik, "Support Vector Networks", Machine Learning, 20, 1995, pp.273-297.
[10] C. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition", Data Mining and Knowledge Discovery, Vol.2, No.2, 1998, pp.121-167.
[11] Y. Q. Chen, X. S. Zhou and T. S. Huang, "One-Class SVM for Learning in Image Retrieval", IEEE Conf. on Image Processing, 2001, pp.34-37.
[12] D. Tax and R. P. W. Duin, "Data Domain Description using Support Vectors", Proc. of European Symposium on Artificial Neural Networks, 1999, pp.251-256.
[13] B. Scholkopf, R. C. Williamson, A. Smola and J. S. Tayler, "SV Estimation of a Distribution's Support", Advances in Neural Information Processing Systems 12, 2000, pp.582-588.
[14] K. Lam and H. Yan, "Locating and extracting the eye in human face images", Pattern Recognition, Vol.29, No.5, pp.771-779, 1996.
[15] R. Chellappa, C. L. Wilson and S. Sirhey, "Human and Machine Recognition of Faces: a Survey", Proceedings of the IEEE, Vol.83, No.5, 1995, pp.705-740.

[16] B. Moghaddam and A. Pentland, "Probabilistic Visual Learning for Object Representation", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.19, No.7, 1997, pp.696-710.

Portable 512 x 512 Grayscale Optical Correlator

Tien-Hsin Chao, Jay Hanan, Hanying Zhou, and George Reyes
Jet Propulsion Laboratory
4800 Oak Grove Drive, Pasadena CA, 91109

Key Words: 512 x 512 Grayscale Optical Correlator, portable systems, CMOS photodetector array with FPGA peak detection processor, automatic target detection

ABSTRACT

JPL is developing a portable 512 x 512 Grayscale Optical Correlator (GOC) system for target data mining and identification applications. This GOC system will utilized a pair of 512 x 512 Ferroelectric Liquid Crystal Spatial Light Modulator (FLCSLM) to achieve 1000 frames/sec data throughput. Primary system design issues including: optics design to achieve compact system volume with fine tuning capability, photodetector array with onboard post-processing for peak detection and target identification. These issues and corresponding solutions will be discussed.

1. GRAYSCALE OPTICAL CORRELATOR TECHNOLOGY OVERVIEW

During the past few years, Jet Propulsion Laboratory (JPL) has developed and demonstrated a matchbox-size compact grayscale optical correlator (GOC) [1-3]. Unique features possessed by JPL developed GOC include: a grayscale input SLM to accommodate direct interface with the input imaging sensor, a real-valued (bipolar-amplitude) filter spatial light modulator to enable use of a MACH (Maximum Average Correlation Height) composite correlation filter algorithm [4-5], compact system architecture due to the use of a pair of 512 x 512 SLM with 7-μm pixel pitch, the smallest developed to date. This GOC architecture has greatly reduced the system complexity by: 1) removing the need of preprocessing (binarization) of the input, and 2) the use of 1 MACH composite filter to greatly reduce the number of filter templates.

The JPL developed 512 x 512 GOC breadboard consists of a diode laser (640 nm), a pair of FLC SLM (512 pixel x 512 pixel, 1000 frames/sec, 7-μm pixel pitch) developed by Boulder Nonlinear Systems Inc. under the NASA SBIR sponsorship, Fourier optics, and a analog CCD photodetector array. The optics head of this GOC breadboard is very compact that only measures 2" x 2" x 1", the size of a matchbox. Figure 1. shows photos. the 512 x 512 FLC SLM and the GOC breadboard.

Figure 1. A photo of a 7-μm pixel pitch, 512 x 512 FLC SLM is shown in 3(a), an arrow pointing to the SLM with the aperture size of 3.58 mm x 3.58 mm. A photo of a 2" x 2" x 1", matchbox-sized GOC is shown in 3(b)

Optical Pattern Recognition XV, edited by David P. Casasent, Tien-Hsin Chao, Proc. of SPIE Vol. 5437 (SPIE, Bellingham, WA, 2004) · 0277-786X/04/$15 · doi: 10.1117/12.548059

The MACH algorithm was optically implemented with the FLCSLM modulated at real-valued mode for automatic target recognition applications. . Merits of the MACH correlation include: a MACH filter is desired to generate an output correlation plane that has a high-peak response with respect to all input images from the same class; The filter maximizes the height of the mean correlation peak relative to the expected distortion (e.g. scale, rotation, perspective angle, etc.); The superior performance of MACH filter is attributed to its unique capability of reducing the filter's sensitivity to distortions and the removal of hard constraints on the peak that permits the optimization of performance criterion. Originally, the MACH filter is of complex-value. JPL has modified it into real-value during optical implementation to accommodate the modulation characteristics of the real-axis FLC SLM.

In conjunction with the use of a MACH filter, the detection of correlation peak relies on the calculation of PSR (peak-to-sidelobe ratio) value, which is defined as: PSR= (peak-μ)/σ, where μ is the mean and σ is the standard deviation over a limited area defined as PSR area. The whole process begins with a search for pixel of maximum intensity (over the entire correlation output). Next find peak intensity (maximum or sum over a small peak area, e.g., 2~4 pixel width). Calculate the mean and standard deviation of the PSR area (usually 16~40 pixel width) that surrounds but excludes the peak area. Calculate the PSR and then nullify the whole PSR area associated with the current peak. This prepares for the next peak search and calculation. Repeat above process until the number of peaks reaches a predefined one (e.g., 10 ~20). Now from all these detected peaks, choose only those whose PSR (and peak intensity) exceeds predefined values.

The PSR peak detection computation was achieved by using the CPU of the post PC. Due to the limit of the state-of-the-art CPU throughput rate, the peak-detection post-processing rate is lower than the 1000 frames/sec and became the processing bottleneck.

2. PRIMARY SYSTEM ISSUES TOWARD BUILDING A PORTABLE 512 X 512 GOC

The following systems issues arise during the design of a portable 512 x 512 GOC: a) compact system layout, b) photodetector array selection for correlation signals detection and c) onboard hardware development for correlation peak detection and target identification.

2.1 COMPACT SYSTEM LAYOUT

The most compact correlator system layout will be achieved by using the traditional 4-f correlator architecture. However, this approach will require precise lens design to accommodate the system parameters such as the laser wavelength, the pixel size match between those of the SLM and the photodetector array. Moreover, the custom lens will be expensive to fabricate.

We would introduce new features into the optics design and achieve compact system layout. For example, the input SLM will be placed behind the Fourier transform lens. It is well known that by doing so, the scale of the Fourier transform of the input image will be linearly proportional to the distance between the input SLM and Fourier transform plane (i.e. back focal plane of the FT lens). The primary adevantage of the approach used in our GOC design over that of the conventional 4-f architecture is the great ease in the stringent precision requirement of the FT lens. As illustrated in equation 1, the focal length is dependent upon the wavelength of the diode laser. This wavelength may vary from device to device within a few nanometers. Thus with the 4-f architecture, very precisive sorting of laser diode has to be accommodate the custom made focal length. In order to accommodate this variation from 4-f design, the inverse Fourier transform lens also have to be modified. A pair of spatially separated lenses have to be utilized to ensure the shortest system light path and precise magnigication to match the input SLM to the photodetector array.

Detail opitcs design of this compact GOC is still under way. Figures 2 shows the rough schematic sketch of the system.

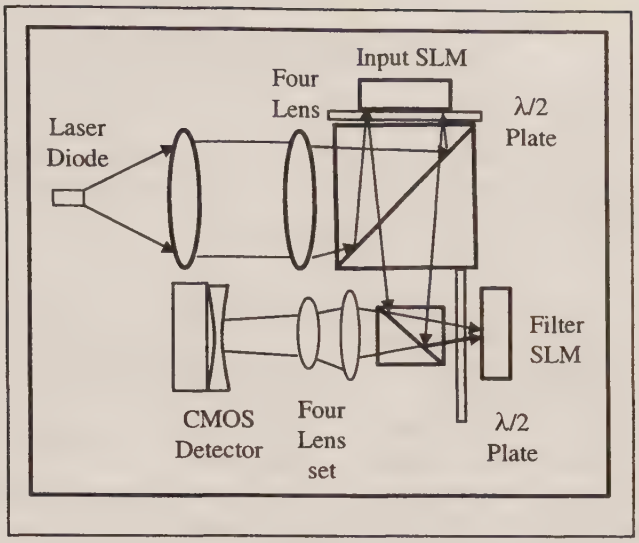


Figure 2. Schematic diagram of a Grayscale Optical Correlator using a pair of reflective type 512 x 512 FLC SLM.

2.2 PHOTODETECTOR ARRAY

In order to match the high frame rate of the FLC SLM, a candidate photodetector array with high-sensitivity, high-speed and digital I/O has to be selected. State-of-the-art digital Charged Couple Device (CCD) such as the Dalsa CCD can only accommodate a modest 256 x 256 size with the frame rate up to 1000 frame/sec. Therefore, the CMOS sensors are being examined as the photodetector array candidate. There are many vendors that provide the CMOS SLM (e.g. Filfactory, Photobit, etc.). We have identified a CMOS sensor with 1024 pixel x 1024 pixel resolution and a maximum speed of 500 frames/sec (when working at full frame size. The speed can be proportionally increased with smaller input frame size). In operation, we would only plan to utilize a quarter of the CMOS sensor area (512 pixel x 512 pixel) and drive the speed to 1000 frames/sec.

2.3 PEAK DETECTION PROCESSING USING ONBOARD HARDWARE

To date, correlation peak detection using post-processing routines such as PSR (peak-to-side-lobe ratio) and PCE (peak correlation energy) has been conducted using the CPU of the host computer. However, as described in the previous section, PSR is a convolution operation. To convolute a 5 x 5 window with the entire 512 x 512 output plane at 1000 frames/sec is beyond the computation capability of the Intel Pentium 4 CPU.

We are currently developing the onboard peak detection hardware using the MaxRevolution frame grabber/FPGA from Datacube Inc. This FPGA approach will enables the removal the system throughput rate bottleneck in peak detection post-processing. It will also enable the use of more robust peak detection post-processing algorithms for further reduction of the false detection rate.

3. SUMMARY

We have briefly introduced the recent JPL effort of developing a portable 512 x 512 GOC with high speed and high resolution. We are currently working on resolving three system issues limiting the

performance of the previous GOC system design. Progress report will be presented during the conference.

4. AKNOWLEDGEMENT

The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. The authors wish to express their sincere thanks to Dr. Geoffrey helpful technical discussions.

5. REFERENCES

1. Tien-Hsin Chao, Hanying Zhou and George Reyes "Grayscale Optical Correlator for Real-time Onboard ATR", Proc of SPIE, Volume 4387 93 (April, 2001).
2. Hanying Zhou and Tien-Hsin Chao, "MACH filter synthesizing for detecting targets in cluttered environment for gray-scale optical correlator," Proceeding of SPIE Vol. 3715, 1999.
3. Tien-Hsin Chao, George Reyes, and Hanying Zhou, "Automatic Target Recognition Field Demo Using a Grayscale Optical correlator ," Proc of SPIE, Vol.3715, pp. 399-406, (1999).
4. A. Mahalanobis, B. V. K. Vijaya Kumar, S. Song, S.R.F. Sims and J. F. Epperson, "Unconstrained correlation filters," *Appl. Opt*, Vol. 33, pp. 3751-3759, (1994).
5. "Real-time Automatic Target Recognition Using a Compact 5 12 x 512 Grayscale Optical Correlator" by Tien-Hsin Chao, Hanying Zhou, George Reyes, and Jay Hanan, presented at SPIE 2003 Optical Pattern Recognition Conference, April 2003

Optical 3D watermarking and authentication by correlation techniques

Sherif Kishk and Bahram Javidi
University of Connecticut, Electrical and Computer Eng. Dept, Storrs, CT, 06269-2157

ABSTRACT

We present an optical method for information watermarking of 3D objects using digital holography and authenticating the hidden image. A hidden image is embedded using double phase encoding in a phase shift digital hologram of the 3D object. The watermarked hologram is decoded to reconstruct the hidden image and the 3D object. We use either the entire hologram or a part of it to decode the hidden image. The recovered hidden image is authenticated using non-linear correlation. Experiments are presented to illustrate the ability to recover both the 3D object and the decoded hidden image.

Keywords: Optical Data Processing, Holography, Three-Dimensional Image Processing, Information Processing

1. INTRODUCTION

Information watermarking[1] can be considered as a method for protecting data from unauthorized distribution. There are vast and important applications of information hiding techniques[1-5]. Double phase encoding[6] can be used to watermark images and has robust properties[7,9] to noise and distortion. The vast majority of digital watermarking work has been reported for 1D signals or 2D images[4]. In this paper we describe a method for watermarking of 3D objects[10,11] using phase shift digital holography[12-16]. Figure 1 shows the phase-shifting interferometic optical system used to record the digital hologram. The laser beam is split into an object beam and a reference beam. We denote the complex reference beam at the CCD plane as,

$$R(x,y;\alpha)=A_R\exp[i(\phi_R+\alpha)] \tag{1}$$

where ϕ_R is the reference beam phase after the second phase retarder when the fast axes are aligned with the polarization direction in both phase retartdation plates. A_R is the amplitude of the reference wave, and α is a phase that can be controlled using the phase retarders. The complex diffraction field of the 3D object at the CCD plane can be presented as:

$$H_O(x,y)=A_O(x,y)\exp[i\phi_O(x,y)] \tag{2}$$

where A_O, and ϕ_O are the 3D object diffraction amplitude and phase at the CCD plane. The phase-shifting interferometer relies on using two phase retardation plates with retardation of $\lambda/2$ and $\lambda/4$, where λ is the laser wavelength. We control the fast and slow axis of the phase retarders to achieve four different intensities with $\alpha_1=0$, $\alpha_2=-\pi/2$, $\alpha_3=-\pi$, $\alpha_4=-3\pi/2$. This is called the four-step method. The Intensity pattern recorded by the CCD is given by

$$I(x,y;\alpha)=\left|H_O(x,y)+R(x,y;\alpha)\right|^2=$$

$$[A_O(x,y)]^2+[A_R(x,y)]^2+2\times A_R(x,y)\times A_O(x,y)\times\cos[\phi_O-\phi_R-\alpha] \tag{3}$$

Optical Pattern Recognition XV, edited by David P. Casasent, Tien-Hsin Chao, Proc. of SPIE Vol. 5437 (SPIE, Bellingham, WA, 2004) · 0277-786X/04/$15 · doi: 10.1117/12.548060

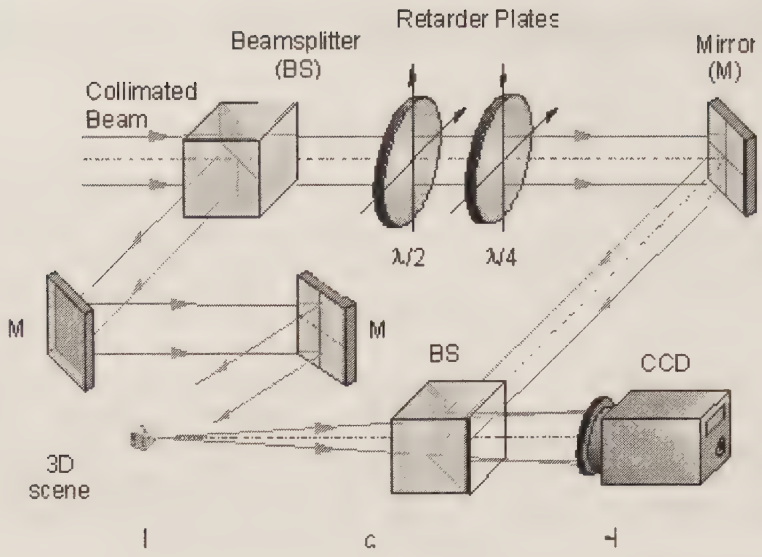


Figure (1) Phase-shifting interferometer for recording digital hologram of a 3D object.

Let us denote the complex wave of interest at the CCD plane by:

$$H_C(x,y) = A_C(x,y)\exp[i\phi_C(x,y)] \tag{4}$$

$H_c(x,y)$ can be used to reconstruct the 3D object . The magnitude of the complex wave at the CCD plane can be calculated using the four step method according to the following formula:

$$A_C = A_O(x,y)A_R = \frac{1}{4}\left\{[I(x,y;0) - I(x,y;-\pi)]^2 + \left[I\left(x,y;-\frac{3\pi}{2}\right) - I\left(x,y;-\frac{\pi}{2}\right)\right]^2\right\}^{1/2} \tag{5}$$

The phase of $H_c(x,y)$ can be calculated from:

$$\phi_C = \arctan\left[\frac{I\left(x,y;-\frac{3\pi}{2}\right) - I\left(x,y;\frac{\pi}{2}\right)}{I(x,y;0) - I(x,y;-\pi)}\right] \tag{6}$$

The calculated amplitude and phase are stored in a digital computer.

Double phase encoding redistributes the energy of the input image such that the encoding data is white Gaussian noise-like. Let the input to the double phase encoder be a complex signal $H(x,y)$. $b_1(x,y)$ and $b_2(x,y)$ are two statistically independent uniformly distributed random variables from 0 to 1. (x,y) are the spatial coordinates. Let $\psi_1(x,y)=exp[j2\pi b_1(x,y)]$ and $\psi_2(\xi,\gamma)=exp[j2\pi b_1(\xi,\gamma)]$ be two phase functions. The double phase encoding signal, $H_d(x,y)$ is given by:

$$H_d(x,y) = \{H(x,y)\psi_1(x,y)\} \otimes IFT[\psi_2(\xi,\gamma)] \tag{7}$$

where (ξ,η) are the coordinates in the Fourier plane and *IFT* stands for inverse Fourier transform. And the symbol $\otimes$ stands for convolution. The decoding process is straightforward.
Using an analysis similar to [8] it can be shown that the encoded double phase encoded complex signal is a white Gaussian noise having a variance given by:

$$\sigma^2 = \frac{1}{N.M}\sum_{y=0}^{M-1}\sum_{x=0}^{N-1} H(x,y)H^*(x,y) \tag{8}$$

where *N,M* are the size of the input signal in the spatial coordinates, and * is the complex conjugate.

Non-linear correlation has the advantage of controlling the balance between distortion tolerance and discrimination by changing the non-linearity coefficient[17]. we perform non-linear correlation between the reference object and the estimates scene. Let the reference object be denoted as *r(x,y)* and the reconstructed input scene be denoted as *s(x,y)*. The correlation output is given by:

$$C = \left| IFT\left[|S|^k \exp(j\phi_S) |R|^k \exp(-j\phi_S) \right] \right| \tag{9}$$

where $|S|$ and $|R|$ are the absolute value of Fourier transform of *s(x,y)* and *r(x,y)*, respectively. ϕ_S and ϕ_R are the phases of the Fourier transform of *s(x,y)* and *r(x,y)*, respectively. *IFT* stands for the inverse Fourier transform, and *k* is the non-linearity that is between 1 and 0.

2. THE PROPOSED SYSTEM

Let $f(x,y)$ be the image to be hidden, (x, y) are the spatial coordinates, and (v, w) are the coordinates in the Fourier domain. $f(x,y)$ is positive and real, and $b(v,w)$ be two independent white sequences, uniformly distributed from zero to one. The double phase encoded image, $\psi(x,y)$, is given by:

$$\psi(x,y) = \{f(x,y)\exp[j2\pi p(x,y)]\} \otimes h(x,y), \tag{10}$$

Where $h(x,y)$ is the impulse response of $H(v,w) = \exp[j2\pi b(v,w)]$, and the symbol $\otimes$ stands for convolution. The watermarked hologram, H_W, is given by

$$H_W(x,y) = H_C(x,y) + a\psi(x,y) \tag{11}$$

where a is an arbitrary constant.
Both the real and imagery part of $\psi(x,y)$ are a Gaussian white process with a variance

$$\sigma^2 = \frac{1}{2\times N\times M}\sum_{x=1}^{N}\sum_{y=1}^{M} f(x,y)^2 \tag{12}$$

where *N* and *M* are the number of pixels of the encoded discrete image. It is straight forward to show that the amplitude of $\psi(x,y)$ has a Raleigh distribution with variance σ^2 and the phase has a uniform distribution. So, the amplitude of the transmitted hologram is distorted by additive Raleigh noise, and the phase of the transmitted hologram is distorted by additive uniformly distributed noise. To recover the hidden image the Fourier transform of the watermarked hologram is multiplied by $\exp[-j2\pi b(v,w)]$, this product is then inverse Fourier transformed, which produces $f(x,y)\exp[j2\pi p(x,y)]$. Finally, $f(x,y)$ can be recovered by multiplying it by $\exp[-j2\pi p(x,y)]$, or by just computing the magnitude of $f(x,y)\exp[j2\pi p(x,y)]$ if $f(x,y)$ is real and positive. The recovered hidden image will have the form:

$$\tilde{f}(x,y) = af(x,y) + IFT\{\hat{H}_C(v,w)\exp[-j2\pi b(v,w)]\}\exp[-j2\pi p(x,y)] \tag{13}$$

where $\hat{H}_C(v,w)$ is the Fourier transform of $H_C(x,y)$. It can be shown that the noise added to the decoded hidden image can be represented by additive white Gaussian noise[8].

We use the normalized mean square error, MSE, as a metric to measure the error in the recovered hidden image and the reconstructed 3D object. The normalized MSE between two images *f(x,y)* and $f_r(x,y)$ can be defined as:

$$MSE = E\left\{\sum_{x=0}^{N-1}\sum_{y=0}^{M-1}\left(|f(x,y) - f_r(x,y)|^2\right)\right\} \Big/ \sum_{x=0}^{N-1}\sum_{y=0}^{M-1}\left(|f(x,y)|^2\right) \tag{14}$$

where E{.} denotes expected value.

3. EXPERIMENTAL RESULTS AND SIMULATIONS

We perform a set of experiments to evaluate the proposed 3D object watermarking under several conditions. We used the minimum mean square error metric to measure the quality of the recovered 3D objects. We generate the host hologram optically using phase shifting holography. The generated hologram is saved digitally using a high resolution CCD.

Figure 2 illustrates the optically generated phase shift digital hologram and the watermarked hologram of the 3D object.

(a) (b)

Figure (2) (a) Phase shifting digital hologram of a 3D object. (b) Watermarked digital hologram of the same 3D object.

Figure 3-a shows the hidden image that will be used as a watermark. Figure3-b presents the reconstructed 3D object without adding watermark to the digital hologram. Figure 3-c illustrates the recovered hidden image that has a normalized MSE of .02. Figure 3-d shows the reconstructed 3D object from the watermarked hologram that has a normalized MSE of .05. Figure 3-e shows the non-linear filter output for a non-linearity of 0.3. As the figure illustrates the correlator was able to authenticate the hidden image.

We can reconstruct the 3D object from different perspectives by using different windows of the recorded digital hologram. A windowed hologram is a fraction of the total hologram. Thus its digital processing is faster.

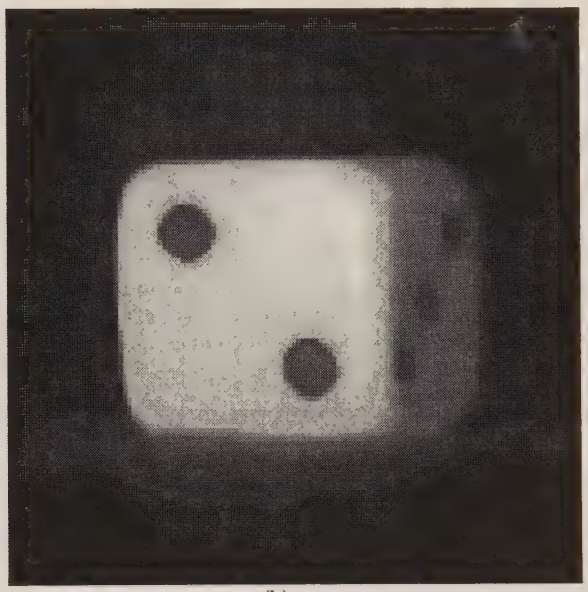
In the next experiment we use a window to reconstruct the digital hologram and use the same window to decode the hidden image. In other words, the hidden image is added to the whole hologram, and we use a part of the hologram to decode both the hidden image and the 3D object. In the following experiment the window size is equal to 25% of the total hologram size and located in the upper left corner of the hologram. This is equivalent to occluding 75% of the encoded hidden image. Figure 4-a shows the reconstructed 3D object using the un-watermarked hologram. Figure 4-b

illustrates the decoded hidden image using the window. The decoded image has a normalized MSE of .07. Figure 4-c presents the reconstructed 3D object from the watermarked hologram using the window. The reconstructed 3D object has a normalized MSE of .075. The next experiment illustrates the ability of the algorithm to recover the hidden image when using only the real part of the hologram. This can be used as a method to reduce the transmission bandwidth. Figure 5-a shows the reconstructed 3D host object using the real part of the watermarked hologram. Figure 5-b shows the decoded hidden image and the corresponding non-linear correlation output when using only the real part of the watermarked hologram. The following experiment tests the use of only 1/16 of the real part transmitted watermarked hologram, which is equivalent to % 98.6 occlusion of the transmitted watermark. Figure 6 shows the reconstructed 3D host object, the recovered hidden image and the corresponding correlator output. The last experiment illustrates the system output against blind detection. Figure 7 shows the decoded hidden image and the corresponding correlator output when wrong codes are used to recover the hidden image.

4. CONCLUSION

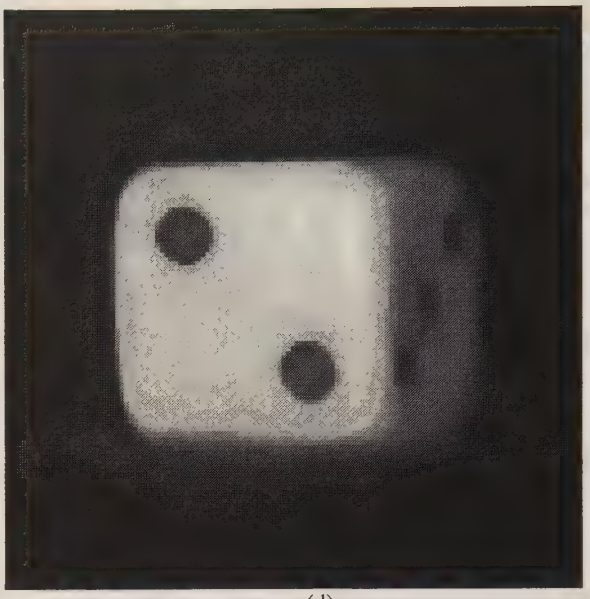
We have presented an optical system for watermarking 3D objects using digital holography. Non-linear correlation is used to authenticate the watermark. The proposed system watermarks the 3D object digital hologram using double phase encoding to encode hidden images. The experiments show that both the 3D object and the hidden watermark can be reconstructed. Also, experiments have shown that we can recover the hidden image from a windowed watermarked hologram to increase the processing speed and reconstruct the 3D object from a different view angle.

(a)

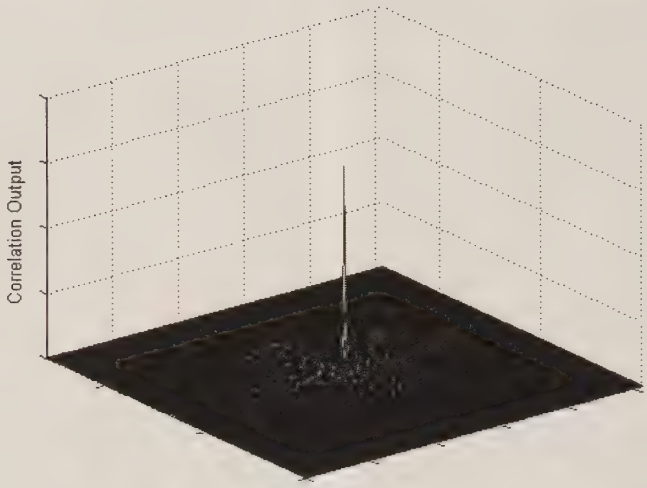


(b)

(c)

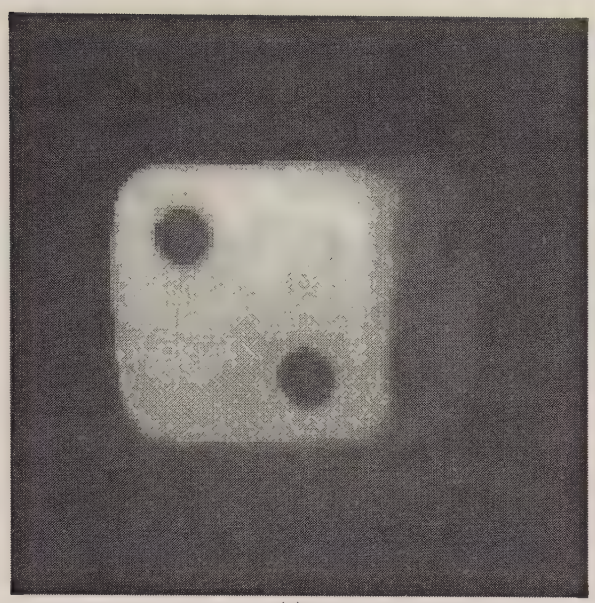


(d)



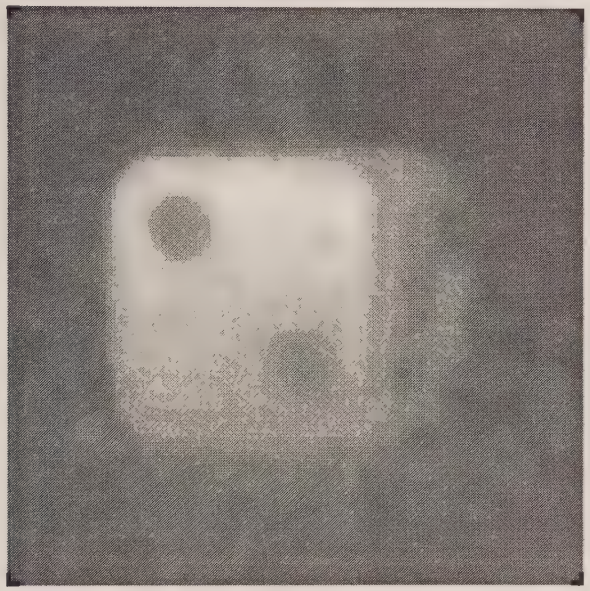
(e)

Figure (3) (a) Image to be hidden (b) Reconstructed 3D object without watermarking (c) Recovered hidden image (d) Reconstructed 3D object using watermarked hologram (e) The non-linear correlation output for detecting the hidden image.

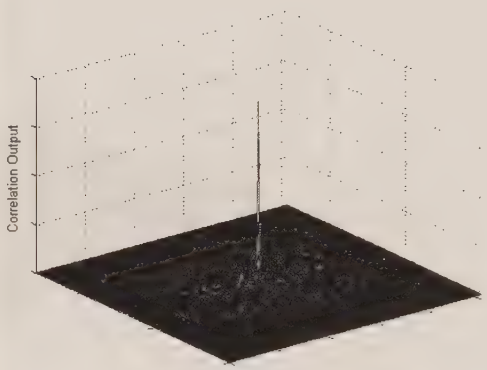


(a)

(b)

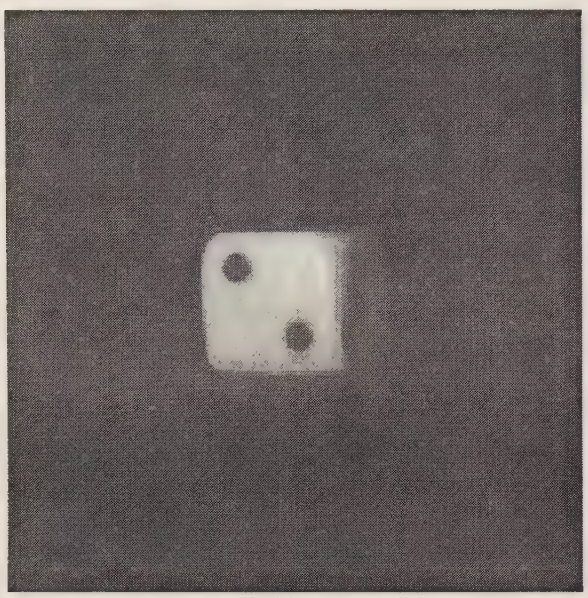


(c)



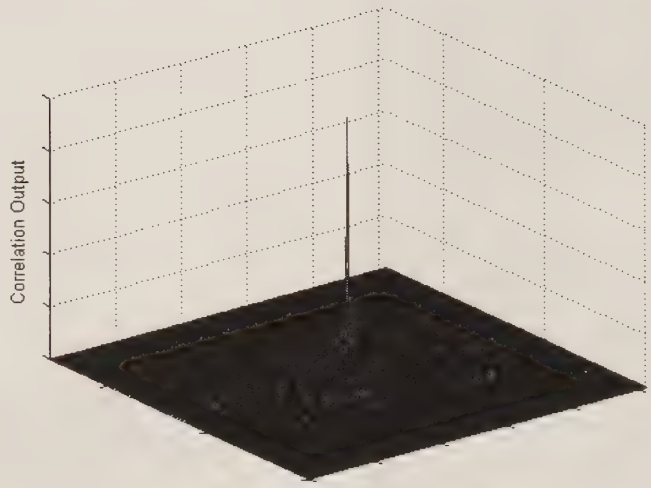
(d)

Figure (4) Watermarking and verification using windowed digital holography
(a) Reconstructed 3D object without watermarking (b) Recovered hidden image (c) Reconstructed 3D object using the watermarked hologram.



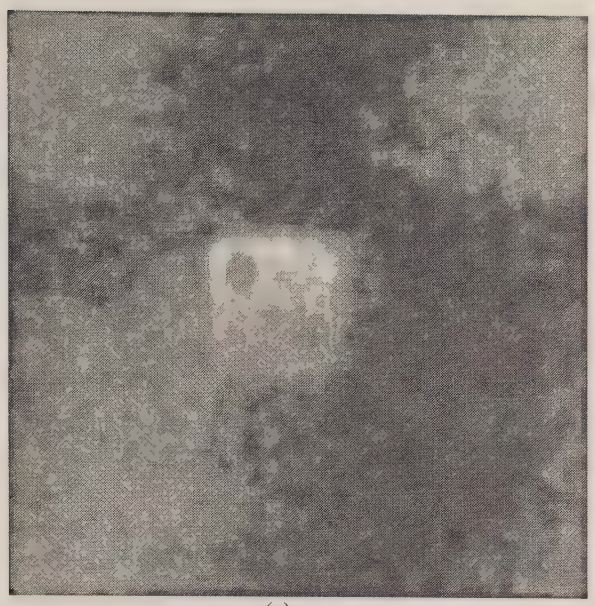
(a)

(b)



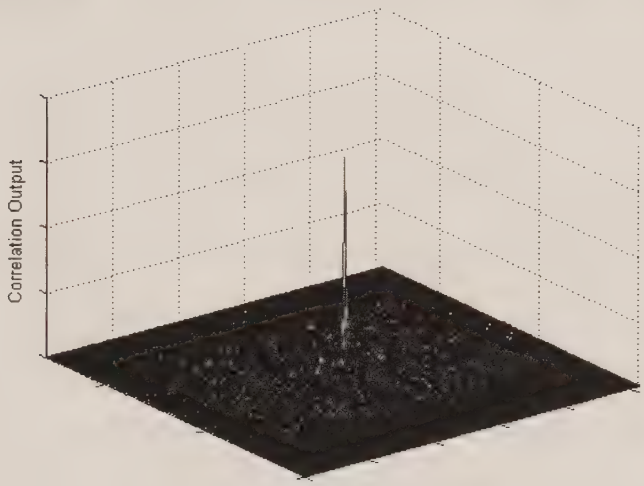
(c)

Figure (5) Watermarking and verification using the real part of the watermarked digital hologram
(a) Reconstructed 3D object from the real part of the watermarked hologram (b) The recovered hidden image (c) The corresponding non-linear correlator output.



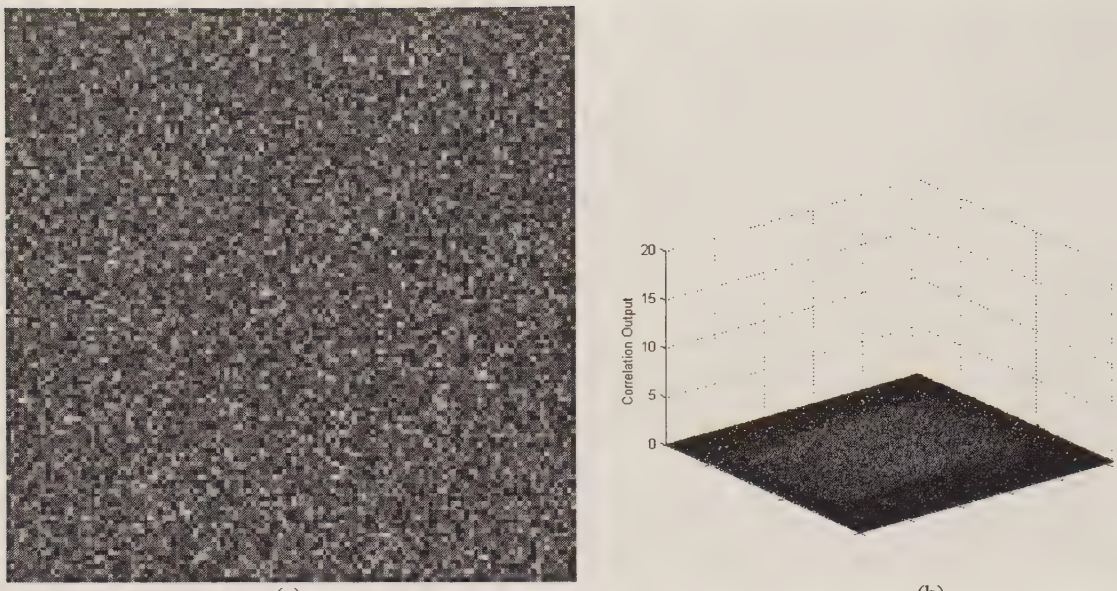
(a)

(b)



(c)

Figure (6) Watermarking and verification using 3.125% of the watermarked hologram and the real part only (a) Reconstructed 3D (b) The recovered hidden image (c) The corresponding non-linear correlator output.



(a) (b)

Figure(7) Blind detection (a) The recovered hidden image using Incorrect codes (b) The corresponding non-linear correlation output.

REFERENCES

1- N. F. Johnson, Z. Duric, and S. Jajodia, *Information Hiding : Steganography and Watermarking - Attacks and Countermeasures (Advances in Information Security, Volume 1,* Kluwer Academic 2001).

2- W. Bender, D. Gruhl, N. Morimoto and L. Lu, "Techniques for data hiding," IBM Systems Journal **35**, 313-336 (1996).

3- Joseph Rosen,Bahram Javidi," Hiding images in Halftone Pictures," Appl. Opt., **40**, 3346 (2001)

4- G. C. Langelaar, I. Setyawan and, R. L. Lagendijk, "Watermarking Digital Image and Video Data. A state of the Art overview," IEEE Signal Processing Magazine **17**, 20-46 (2000).

5- C. Hosinger and M. Rabbani, "Data Embedding using Phase Dispersion," The International Conference on Information Technology: Coding and Computing (ITCC 2000), Las Vegas, Nev. 27-29 March (2000).

6- P. Refregier and B. Javidi, "Optical image encryption using input plane and Fourier plane random encoding," Opt. Let., **20**, 767-769 (1995).

7- R. K. Wang, I. A. Watson, and C. Chatwin, "Random phase encoding for optical secutity," Optical Engineering, **35**, 2464-2460 (1996).

8- S. Kishk and B.Javidi, "Information hiding technique with double phase encoding," Applied Optics **41**, 5462-5470 (2002).

9- F. Goudail, F. Bollaro, B. Javidi, and P. Refregier, "Influence of perturbation in a double phase-encoding system," J. Opt. Soc. Am. A, **15**, 2629-2638 (1998).

10- S. Kishk and B. Javidi "Watermarking of a 3D Object Using Digital Holography," Opt. Let., **28**, 167-169 (2003).

11- S. Kishk and B. Javidi, "3D object watermarking by a 3D object," Optics Express 11, 874-888 (2003).

12- H. J. Caulfield, *Handbook of Optical Holography*, Academic, London, 1979.

13- I Yamaguchi, and T Zhang, "Phase-Shifting Digital Holography," Opt. Lett. **22**, 1268-1270 (1997)

14- J. H. Bruninig, D. R. Herriott, J. E. Gallagher, D. P. Rosenfeld, A. D. White, and D. J. Brangaccio, "Digital Wavefront measuring interferometer for Testing Optical Surfaces and Lenses," Appl. Opt. **13**, 2693 (1974).

15- J. Schwider, "Advanced Evaluation Techniques in Interferometry," in *Progress in Optics*, E. Wolf, ed (North Holland, Amesterdam) **XXVIII**, 271-359 (1990)

16- T. Naughton, Y. Frauel, B. Javidi, and E. Tajahuerce," Compression of digital holograms for three-dimensional object reconstruction and recognition," Appl. Optics, **11**, 4124-4132 (2002).

17- B. Javidi, ''Nonlinear joint power spectrum based optical correlators,'' Appl. Opt. **28**, 2358–2367 (1989).

Efficient method for extracting object recognition metric from correlation filter output

B.V.K. Vijaya Kumar, and Jason Thornton
Dept. of Electrical and Computer Engineering
Carnegie Mellon University, Pittsburgh, PA 15213

ABSTRACT

Correlation filters can be very effective for object recognition. However, these filters may become too computationally expensive when applied to large images, or large numbers of images, because they require Fourier transforms between the spatial and frequency domains. This paper makes the simplifying assumption of single object recognition and presents an algorithm designed to reduce complexity of computation and/or storage. The algorithm derives a frequency domain match metric as opposed to the standard approach of using the spatial correlation plane. The performance of the efficient algorithm is compared to that of the standard correlation filter algorithm, for both accuracy and computational requirements.

Keywords: Correlation filters, Biometrics, Efficiency, Peak-to-sidelobe ratio

1. INTRODUCTION

Correlation filters [1] have been investigated for use in automatic target recognition (ATR) applications [2] and for biometric applications [3] among others. In these and other similar applications, a correlation filter is designed from a set of training images during the enrollment stage and during the verification stage, a test image is processed by the correlation filter to yield to a correlation output. For a well-designed filter, the correlation output should exhibit a sharp peak for an authentic input image and no such peak for an impostor input image. Since the peak sharpness is supposed to reflect the match of the input image to the authentic image, a metric such as the peak-to-sidelobe ratio (PSR) is usually computed from the correlation output and that PSR is compared to a pre-selected threshold to determine whether the input image is of an authentic or an impostor.

In ATR applications, the input scene can contain multiple targets (some from the desired class and others from clutter class), possibly anywhere in the scene. Thus, correlation filters must be able to handle multiple targets and also must exhibit shift-invariance in the sense that correlation peak can be anywhere in the output plane. However, in many biometric recognition applications, the input scene contains only one object and that too is usually more or less centered. For example, consider the use of face verification for accessing a secure area. There will only be one face in the scene and the subject (even an impostor) will do his/her best to cooperate by centering their face. In such applications involving single objects that are reasonably centered, we may be able to assume that the correlation peak should be close to the origin and that the correlation output contains only a single peak. These assumptions can lead to a method to efficiently extract a recognition metric (e.g., peak-to-sidelobe ratio, peak-to-correlation energy) from the correlation output.

The rest of this paper is organized as follows. Section 2 provides some background for correlation filters. Section 3 discusses how object recognition metrics can be computed in the frequency domain. In Section 4,

Optical Pattern Recognition XV, edited by David P. Casasent, Tien-Hsin Chao, Proc. of SPIE Vol. 5437 (SPIE, Bellingham,

we analyze the complexity of the proposed method. Section 5 presents some illustrative results and Section 6 contains our conclusions.

2. CORRELATION FILTERS

Image matching can be performed by cross-correlating an input image with a synthesized template (i.e., the inverse Fourier transform of a correlation filter) and analyzing the resulting correlation output. Figure 1 shows schematically how the cross-correlation is obtained using Fourier transforms. The correlation output is searched for peaks, and the heights of these peaks as well as other related metrics (e.g., peak-to-sidelobe ratio or PSR) are used to determine whether the input is an authentic or an impostor.

The most basic correlation filter is the matched filter (MF), which is designed to detect a known reference image corrupted by additive white noise. The MF performs poorly when the reference image appears with distortions (e.g., rotations, scale changes). In the presence of distortions, one MF will be needed for each appearance. Clearly this is computationally impractical. Composite correlation filters or the synthetic discriminant function (SDF) filters introduced by Hester and Casasent [4] address this challenge. The SDF filter can be interpreted as a linear combination of MFs where the combination weights are chosen so that the correlation outputs corresponding to the training images would yield pre-specified values at the origin. For example, the correlation peak values corresponding to the training images of authentics can be set to 1, and the peak values due to the impostor training images can be set to zero. It is hoped that the resulting correlation filter would yield correlation peak values close to 1 for non-training images from the authentic class and correlation peak values close to zero for non-training images from the impostor class.

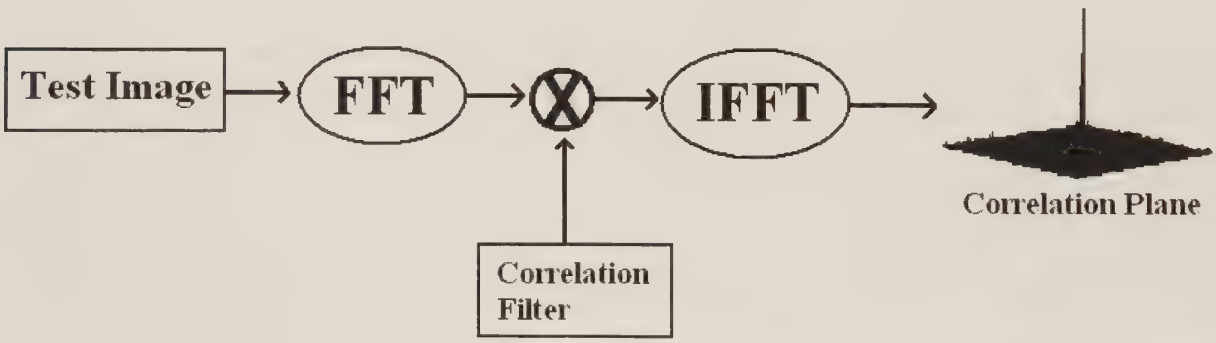


Figure 1 Block diagram of correlation process

Although the original SDF filter produces pre-specified correlation peak values, it also results in large sidelobes. Sometimes these sidelobes are larger than the pre-specified peak values leading to misclassifications. The peak sharpness can be improved by using minimum average correlation energy (MACE) filters [2] designed to produce correlation peaks. More details about correlation filter designs can be found elsewhere [1].

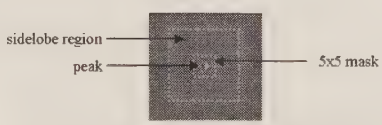


Figure 2: Estimation of the peak-to-sidelobe ratio (PSR)

When a correlation filter is applied to an image, the resulting correlation plane is processed to determine a match metric. The metric in some way quantifies the presence of a sharp correlation peak. A common way to do this is to use the Peak-to-Sidelobe Ratio (PSR), which defines the match metric in terms of the height of the tallest peak and the shape of the correlation plane that surrounds that peak. The estimation of the peak-to-sidelobe ratio (PSR) is illustrated in Fig. 2. First, the correlation peak is located, and a small (e.g., of size 5x5) mask is centered at the peak. The sidelobe region is defined as the annular region between this small mask and a larger (e.g., of size 20x20) square also centered at the peak. The mean and standard deviation of the sidelobe region are computed and used to estimate the PSR as follows.

$$\text{PSR} \triangleq \frac{|\text{peak height} - \text{mean}|^2}{\text{variance of sidelobes}} \tag{1}$$

The recognition or verification decision is made by comparing the estimated PSR to a fixed threshold. If the test image contains the object that the correlation filter was designed to recognize, the correlation plane should have a distinct peak called (the correlation peak) and yield large PSR. If the test image does not contain a member of that object class and no match occurs, then the correlation plane should show no strong peak and the PSR will be small.

The traditional method to compute the match metric is to form the correlation output, locate the peak, compute the mean and the standard deviations in the sidelobe region centered at the peak (as shown in Fig. 2) and compute the PSR according to Eq. (1). However, as can be seen from Fig. 1, producing the correlation output requires that we perform two 2-D fast Fourier transform (FFTs) for every correlation filter. In verification applications, one test image is tested against one filter. However, in identification applications, one test input is tested against several correlation filters requiring the computation of many 2-D FFTs, which can be computationally demanding. In the next section, we discuss how to reduce this complexity by evaluating the peak sharpness metrics directly in the frequency domain, without forming the correlation outputs (i.e., by omitting the second of the two 2-D FFTs).

3. COMPUTING A FREQUENCY DOMAIN MATCH METRIC

The peak sharpness metric can be defined in different forms. For example, the size and shape of the sidelobe region surrounding the peak can take on different values. For the purposes of the algorithm presented in this section, we consider a match metric of a more general form than the PSR in Figure 2. In this form, the peak is compared to the entire correlation plane instead of a limited sidelobe region. Because it is assumed that the correlation plane contains a single peak, the sidelobe region can be extended to the entire plane without the concern of including other correlation peaks.

For a correlation plane of size M by N, let $\{x\}$ represent the values of the correlation plane, so that x_{rs} is the value from row r and column s. Define μ_x as the mean of the correlation plane, i.e.,

$$\mu_x \triangleq \frac{1}{MN} \sum_{r=0}^{M-1} \sum_{s=0}^{N-1} x_{rs} \tag{2}$$

Now we can define the general match metric as follows.

$$\text{match metric} \triangleq \max_{p_1,p_2}\left[\frac{|x_{p_1p_2}-\mu_x|^2}{\frac{1}{MN}\sum_{r=0}^{M-1}\sum_{s=0}^{N-1}|x_{rs}-\mu_x|^2}\right] \tag{3}$$

This metric is equivalent to taking the zero-mean version of the complex plane, then finding the ratio of the peak's squared magnitude to the average squared magnitude. As the correlation peak approaches its ideal shape of a delta function, this match metric increases (as desired).

The correlation plane is the inverse discrete Fourier transform (DFT) of the filtered frequency domain. Some properties of the correlation plane can be inferred from the frequency domain, without actually performing an inverse DFT. This section shows one way to transform the match metric so that it is in terms of the frequency plane values.

First, it is assumed that the peak location can be determined without searching the correlation plane. This simplifying assumption holds when using generally centered images (since the correlation peak occurs at the origin). This is a reasonable constraint for some applications, such as in face verification where the subjects will usually be cooperative and can be expected to align their faces to some crosshairs or other markers. Since there is no need to search for a correlation peak, the match metric can be rewritten as follows.

$$\text{match metric} = \frac{|x_{p_1p_2}-\mu_x|^2}{\frac{1}{MN}\sum_{r=0}^{M-1}\sum_{s=0}^{N-1}|x_{rs}|^2 - |\mu_x|^2} \tag{4}$$

Let the frequency plane values be represented by X_{kl}, the value from row k and column l. Note that the frequency plane values are related to the correlation plane values according to the following DFT relation.

$$X_{kl} = \frac{1}{MN}\sum_{r=0}^{M-1}\sum_{s=0}^{N-1}x_{rs}e^{-jkr(2\pi/M)}e^{-jls(2\pi/N)} \tag{5}$$

Then the origin of the frequency plane is

$$X_{00} = \frac{1}{MN}\sum_{r=0}^{M-1}\sum_{s=0}^{N-1}x_{rs} = \mu_x \tag{6}$$

So the value at the frequency plane origin can be used for the correlation plane mean μ. Also, the double summation in the denominator of Eq. (4) calculates the average energy of the correlation plane. This can be replaced by the energy of the frequency domain as follows.

$$\frac{1}{MN}\sum_{r=0}^{M-1}\sum_{s=0}^{N-1}|x_{rs}|^2 = \sum_{k=0}^{M-1}\sum_{l=0}^{N-1}|X_{kl}|^2 \tag{7}$$

Finally, the peak value $x_{p_1p_2}$ can be calculated with the inverse DFT as a projection of the frequency plane values onto a single complex exponential basis, i.e.,

$$x_{p_1p_2} = \sum_{k=0}^{M-1}\sum_{l=0}^{N-1} X_{kl}\, e^{jkp_1(2\pi/M)} e^{jlp_2(2\pi/N)} \tag{8}$$

Substituting Eqs. (6), (7 and (8) into (4), the match metric becomes the following in terms of strictly frequency domain quantities.

$$\text{match metric} = \frac{\left|\sum_{k=0}^{M-1}\sum_{l=0}^{N-1} X_{kl}\, e^{jkp_1(2\pi/M)} e^{jlp_2(2\pi/N)} - X_{00}\right|^2}{\sum_{k=0}^{M-1}\sum_{l=0}^{N-1}|X_{kl}|^2 - |X_{00}|^2} \tag{9}$$

This match metric is now expressed in terms of frequency domain values and not correlation plane values. Assuming that the peak location (p_1, p_2) is known, all computation can be done without explicitly generating the correlation plane. This saves the computation of one inverse DFT for every filter/image combination. In verification applications, computing the PSR type match metrics directly in the frequency domain saves one inverse DFT for each test image. However, in identification, each test image is compared against many filters (as many as the number of classes in a database) generating many PSRs and final identification is based on the filter leading to the largest PSR. Thus, in identification applications, we save as many inverse DFTs per test image as the number of subjects in the database. Assuming that the pattern is roughly centered leads to further simplification. With the assumption that $p_1 = p_2 = 0$, Eq. (9) simplifies the following convenient form.

$$\text{match metric} = \frac{\left|\sum_{k=0}^{M-1}\sum_{l=0}^{N-1} f_{kl} - f_{00}\right|^2}{\sum_{k=0}^{M-1}\sum_{l=0}^{N-1}|f_{kl}|^2 - |f_{00}|^2} \tag{10}$$

4. COMPLEXITY ANALYSIS

When correlation filters are applied to a test image, the image must first be converted to its frequency domain by a DFT. This transform is implemented using a fast Fourier transform (FFT) algorithm, which has computational complexity $O(NM \log_2 NM)$, where the image is of size M by N. This 2-D frequency domain array is multiplied by the 2-D filter array and the resulting product array is processed to determine a match metric.

Tables 1-3 show the computational complexity of 3 methods for evaluating the application of a correlation filter. The tables do not include the transition of the image to the frequency domain, as this does not have to be repeated for each filter that is applied to the image. The first algorithm is the standard way of generating the correlation plane and applying Equation (3). The second algorithm is the frequency-domain calculation of a match metric, assuming the peak location can be estimated (Eq. (9)). The third algorithm is the frequency-domain calculation of a match metric, assuming a peak location at the origin (Eq. (10)). The tables partition the algorithms into their different steps, and show the number of multiplications/divisions required at each step as well as the O-notation (the asymptotic upper bound on computational complexity, ignoring all constants). Note that the computation of squared magnitude is counted as a single complex multiplication.

Table 1. Standard algorithm

Step	*O-notation*	# of mult./div.
Apply correlation filter with point-by-point multiplication.	$O(MN)$	MN
Generate correlation plane with inverse FFT.	$O(MN \log_2 MN)$	$MN \log_2 MN$
Compute mean / subtract from plane.	$O(1)$	*constant*
Take squared magnitude of plane.	$O(MN)$	MN
Compute average squared magnitude.	$O(1)$	*constant*
Divide peak value by average squared magnitude.	$O(1)$	*constant*
Total	$O(MN \log_2 MN)$	$MN \log_2 MN + 2MN + constant$

Table 2. Frequency domain algorithm

Step	*O-notation*	# of mult./div.
Apply correlation filter with point-by-point multiplication.	$O(MN)$	MN
Compute peak value.[1]	$O(MN)$	MN
Compute $\lvert \text{peak} - \text{mean} \rvert^2$.	$O(1)$	*constant*
Find energy of frequency plane.	$O(MN)$	MN
Find variance of frequency plane.	$O(1)$	*constant*
Divide peak $\lvert \text{peak} - \text{mean} \rvert^2$ by variance of plane.	$O(1)$	*constant*
Total	$O(MN)$	$3MN + constant$

Table 3. Frequency domain (peak at origin) algorithm

Step	*O-notation*	# of mult./div.
Apply correlation filter with point-by-point multiplication.	$O(MN)$	MN
Sum frequencies to get peak.	----	0
Compute $\lvert \text{peak} - \text{mean} \rvert^2$.	$O(1)$	*constant*
Find variance of frequency plane.	$O(MN)$	$MN+1$
Divide peak $\lvert \text{peak} - \text{mean} \rvert^2$ by variance of plane.	$O(1)$	*constant*
Total	$O(MN)$	$2MN + constant$

Algorithm 1 is more complex because it relies on an inverse FFT to get back to the spatial domain. Algorithms 2 and 3 have linear complexity with respect to the total number of image pixels (MN), and should run faster because of reduced number of multiplications/divisions. If the complex exponential basis used in Algorithm 2 is not pre-computed in storage, and has to be computed at run-time, computational

[1] This assumes that the complex exponential basis used to calculate the peak value has been pre-computed and stored.

complexity remains the same but a number of costly calculations are added. Algorithm 3 does not have this requirement, and in any case offers the most efficient method when dealing with centered images.

5. TESTING

For comparison, the regular correlation filter algorithm and a frequency-domain version of the algorithm were implemented and tested. The purpose of these tests was not to evaluate or improve the performance of correlation filters in general, but to compare the relative performance of algorithms (in terms of accuracy and speed).

The images used for testing were derived from the facial expressions database collected in the Advanced MultiMedia Processing (AMP) Laboratory at CMU [5]. The selected subset has been used before in correlation filter experiments, and is well suited for recognition using correlation filters (it does not pose any significant challenges to correlation filters). The dataset consists of 13 classes of cropped face images. Each class contains 74 images of the same person's face with varying expressions. Figure 3 shows three sample images from one subject.

Figure 3 Sample facial expressions images from one subject

Of course, in this dataset there is only one target pattern per input image and thus only one correlation peak. In addition, this dataset accommodates the assumption that the correlation peak is roughly centered around the origin, which is a necessary condition for the third algorithm presented.

For testing, one correlation filter was generated for each class using seven training images from that class. The seven training images were selected at evenly distributed indices in the class, and this number was chosen empirically to achieve reasonable recognition performance. The filters were designed as optimal trade-off synthetic discriminant function (OTSDF) filters [6], with a trade-off parameter of 0.5. These types of correlation filters are designed to balance good discrimination with tolerance of noisy images. After training, 67 images remain from each class for testing. These test images were applied to the correlation filter corresponding to their own class, as well as the correlation filters corresponding to all other classes. Testing was conducted on a centered dataset, so algorithm 3 was tested instead of algorithm 2. After every test image was filtered in its frequency domain, match evaluation was computed in two ways:

- with the standard method of generating a correlation plane (labeled above as algorithm 1)
- with the modified frequency-domain method, assuming centered images (labeled above as algorithm 3)

For the task of face verification, a test image claims to belong to a certain class and the algorithm must either confirm or deny that claim. In this context, the recognition algorithm returns one of two responses: accept, or reject. This is decided by comparing the match score to a threshold and accepting all matches above the threshold. Adjusting the threshold changes the rate of false accepts and the rate of false rejects. This tradeoff is shown in a Receiver Operating Characteristic (ROC) curve, which plots False Accept Rate (FAR) against False Reject Rate (FRR). The Equal Error Rate (EER) is where FAR equals FRR in the trade-off.

Verification performance of the standard algorithm was compared to that of the modified frequency domain algorithm. The two ROC sets are shown in Figs. 4 and 5. For each algorithm, we show the ROC curve of the most difficult subject (of the 13 subjects), the ROC curve of the easiest subject, and the all-class ROC curve which uses a single accept/reject threshold for all face classes put together.

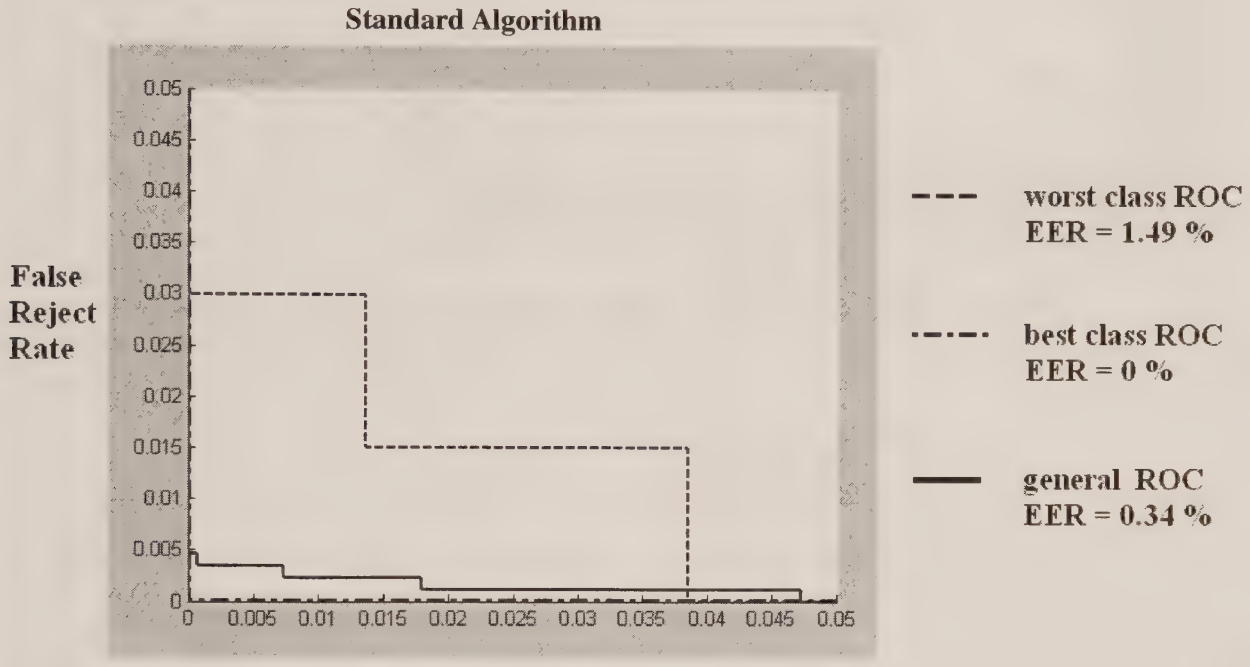


Figure 4. ROC curves for the standard algorithm

The verification performance varies with class, with class 2 of the dataset being the most to difficult to verify. Note that the EER of the frequency domain algorithm is slightly higher than that of the standard algorithm in the worst class ROC. However, the general EER actually improves from 0.34 % to 0.14 %.

To measure computation time for different input dimensions, random complex-valued matrices of different sizes were generated to simulate potential images and correlation filters. The timing trials were run in Matlab on a Pentium 4, 2.26 GHz CPU with 640 MB of RAM. The average time required per image for the standard algorithm to evaluate for match is shown in row 1 of Table 4. The average time required for the frequency domain algorithm is shown in row 2 of Table 4.

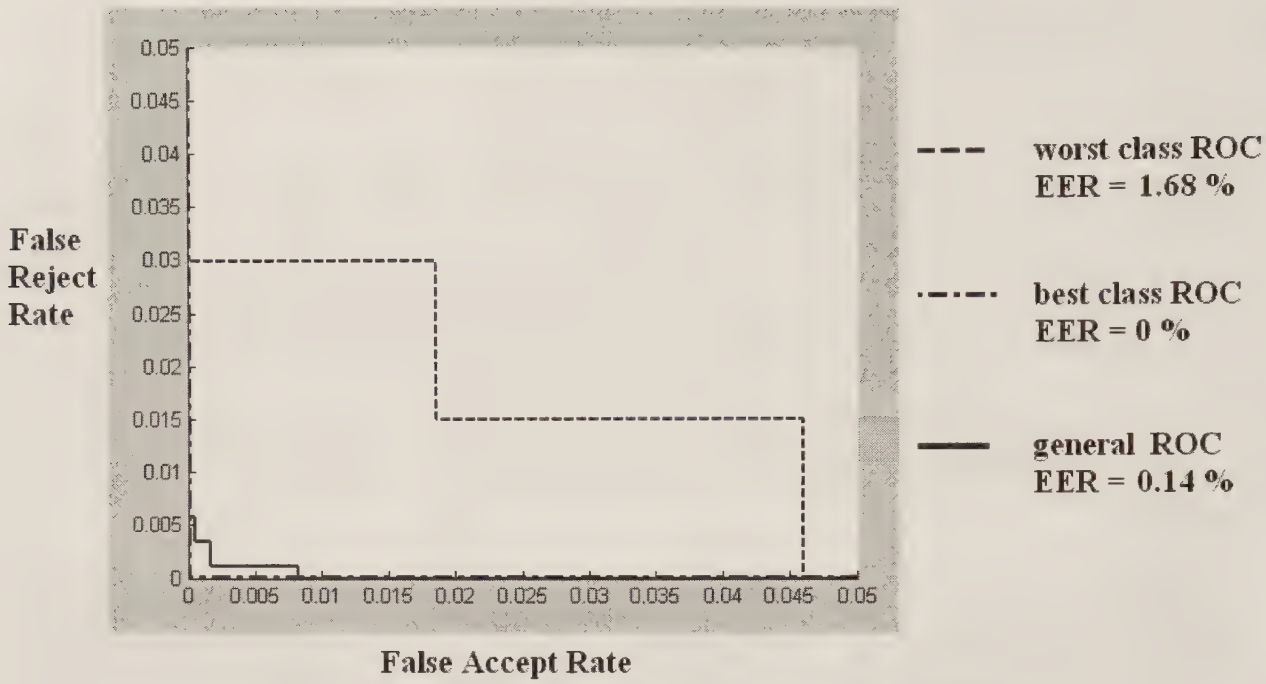


Figure 5. ROC curves for the frequency domain algorithm

Table 4. Computation time comparison

	Image Size					
	50 by 50	100 by 100	200 by 200	400 by 400	800 by 800	1600 by 1600
Standard Algorithm	1.6×10^{-3} s	6.4×10^{-3} s	0.026 s	0.14 s	0.60 s	3.60 s
Frequency-Domain Algorithm	4.7×10^{-4} s	2.2×10^{-3} s	8.1×10^{-3} s	0.040 s	0.13 s	0.72 s

6. DISCUSSION

The standard algorithm and the frequency domain algorithm for evaluating a match appear to yield have comparable verification accuracies. The frequency domain algorithm actually gives a slight improvement in the general ROC curve (one threshold for all classes). This is a surprising result, but plausible as long as the assumption made for the modified algorithm is correct, meaning that the images in the dataset are roughly centered. Because the frequency-domain algorithm does not require the computation of an inverse FFT, it

shows some improvement in execution time. This improvement ranges from a factor of approximately 3 for smaller images to a factor of approximately 5 for larger images. That the advantage increases with image size is consistent with the O-notation complexities of the algorithms.

Acknowledgments

The authors would like to thank the United States Army Research Office (ARO) for partial support of this research. The authors also acknowledge some earlier research along these lines by Krithika Venkataramani of Carnegie Mellon University.

REFERENCES

1. B.V.K. Vijaya Kumar, Tutorial survey of composite filter designs for optical correlators," *Applied Optics*, Vol. 31, pp. 4773-4801, 1992.
2. A. Mahalanobis, B. V. K. Vijaya Kumar and D. Casasent, "Minimum average correlation energy filters," *Applied Optics*, Vol. 26, pp. 3633-3630, 1987.
3. M. Savvides, B.V.K. Vijaya Kumar and P. Khosla, "Face verification using correlation filters," *Proc. Of the Third IEEE Automatic Identification Advanced Technologies*, Tarrytown, NY, March 2002, pp. 56-61.
4. C.F. Hester and D. Casasent, "Multivariant Technique for Multiclass Pattern Recognition," *Appl. Opt., Vol. 19*, 1980, pp.1758-1761.
5. Facial expressions database, AMP Laboratory, Carnegie Mellon University.
6. Ph. Réfrégier, "Optimal trade-off filters for noise robustness, sharpness of the correlation peak, and Horner efficiency," *Optics Letters,* Vol. 16, pp. 829-831, 1991.

Invariant Fringe-adjusted Joint Transform Correlation-based Target Tracking in FLIR Sequences

Chye-Hwa Loo and Mohammad S. Alam*
Department of Electrical and Computer Engineering,
University of South Alabama, Mobile, AL 36688

ABSTRACT

Presented in this paper is a fringe-adjusted joint transform correlator (FJTC) based invariant target tracking of forwar looking infra-red (FLIR) image sequences. The proposed FJTC based tracking approach employed a modified synthetic discriminant function (SDF) concept together with an efficient camera motion compensation technique to accommodate the problem of target signature variation due to in-plane/out-of-plane rotations, scale variations, noise, and bad frames. The proposed technique can track small objects comprising of only a few pixels and is capable of compensating the high ego-motion of the sensor. The robustness of the proposed technique is demonstrated with computer simulation performed on sequences of eal life FLIR imagery taken from an airborne moving platform.

Keyword: FLIR imagery, distortion-invariant target tracking, ego-motion compensation, fringe-adjusted joint transform correlation, FJTC, synthetic discriminant function, SDF.

1. INTRODUCTION

Although a great deal of effort has been expended on detecting objects in visual images, only limited amount of work has been reported on the detection and tracking of targets in infrared images. In general, existing methods on IR images work for a limited number of situations due to the constraints imposed on the solution. In contrast to visual images, thermal images such as forward-looking infrared (FLIR) imagery usually have low resolution (128x128), extremely low signal-to-noise ratio, low contrast, and very small target signature (a few pixels). Moreover, the variation of the target signature, competing background clutter, lack of *a priori* information, high ego-motion of the sensor, and artifacts generated due to the weather conditions complica te the detection and tracking of targets.

To detect targets in FLIR image sequences, majority of the existing algorithms rely on the hot spot technique, which assume that the target radiation is much stronger than the radiation of the background and noise [1,2]. However, in realistic tracking scenarios, targets do not always appear brighter than the background. In contrast, motion detection methods based on motion tracking require background textures to differentiate the target movements [2-4]. Therefore, these techniques cannot effectively detect independently moving objects when the background contrast is low. In addition, motion detection methods yield false alarms caused by moving clutter and cannot effectively track non-moving targets. Similarly, some target tracking algorithms work only for sequences with no sensor ego-motion [5]. Other tracking methods [6,7], which are based on mean-shift tracking rely on the intensity distributions generated from the target region and computes the translation of cent er of the target in the image space. These algorithms [6,7] are more efficient as they do not confine to bright targets and are able to compensate for sensor ego-motion. In general, most of the aforementioned methods are computation intensive and are not suitable for real-time applications.

In this paper, we present a novel approach for real-time target tracking in FLIR imagery in the presence of high global motion as well as various distortions in target features. The proposed approach utilizes fringe-adjusted JTC (FJTC) technique for real-time estimation of target motion. For real-time optical pattern recognition, the JTC technique has been found to be a robust alternative to VanderLugt type correlation. Although a classical JTC preserves the shift-invariant pattern recognition property, it suffers from high sensitivity to scale and rotation variations, large correlation side-lobes, and a high zer -order peak. Various improved versions of JTCs have been developed such as binary JTC [8], amplitude modulated filter (AMF) based JTC [9], and fringe-adjusted filter (FAF) based JTC [10-12]. Among these

*malam@usouthal.edu; phone 1 251 460 6117; fax 1 251 460 6028

Optical Pattern Recognition XV, edited by David P. Casasent, Tien-Hsin Chao, Proc. of SPIE Vol. 5437 (SPIE, Bellingham, WA, 2004) · 0277-786X/04/$15 · doi: 10.1117/12.548061

techniques, the fringe-adjusted JTC (FJTC) has been found to yield better correlation performance while alleviating th problems associated with alternate techniques.

In order to track a target in an input scene, a known reference image and the unknown input scene are introduced side-by-side in the input joint image for FJTC processing. The main theme of this research is target tracking. Accordingly, it is assumed that the initial position and size of the target in the first frame of the FLIR sequence are known. To track a target in a FLIR image sequence, the reference image corresponding to the target to be detected is obtained from a database and the input image i s obtained from the segmented initial and subsequent frames of th sequence. The proposed technique has been tested with real FLIR imagery supplied by Army Missile Command (AMCOM) and has been found to yield excellent tracking performance. This algorithm is capable of tracking both moving and stationary targets while compensating the global motion or ego-motion caused by the sensor. In addition, it can also track low-contrast objects and very small objects involving only a few pixels.

Because FLIR imagery is generally corrupted by noise, a target may appear dissimilar from frame to frame. A frame is considered a bad frame by the tracking algorithm if it is corrupted by noise to the extent that the algorithm fails to track the target. Discarding the bad frame might solve the problem but the selection and determination of whether particular frame is to be discarded requires the setting and application of a criterion or threshold. Selecting a wrong threshold value may cause the tracking algorithm to discard to o many frames and, therefore frame discarding may not lead to a reliable solution. In this paper, we utilized a concept similar to the synthetic discriminant function (SDF) in the FJTC based tracking algorithm that alleviates the effects of noise, 3D distortions, and hazards generated by bad frames yielding significantly improved overall tracking performance.

The remainder of the paper is organized as follows: Section 2 analyzes the mathematical concept of FJTC motion estimation process followed by the analysis of the target model update with SDF in Section 3. In Section 4, we present the design and implementation details associated with the proposed FJTC based tracking. Finally, simulation results for the proposed tracking technique using real life FLIR imagery are presented in Section 6 and conclusions are included in Section 7.

2. FJTC BASED MOTION ESTIMATION

In this section, we briefly discuss how FJTC can be used in motion estimation. Figure 1 shows the architecture for a real-time FJTC system. The reference image and the input scene, captured by the camera C1, are introduced in the input plane by using an SLM such as a liquid crystal television. Assume that the reference image which corresponds to the target to be tracked is denoted b $r(x,y+y_0)$, and the input scene which may contain the target is denoted by $t(x-x_i, y-y_i-y_0)$, where (x_i, y_i) are the arbitrary coordinates of the target in the input scene. If the input scene is also assumed to be corrupted by noise $n(x, y-y_0)$, then the input joint image can be expressed as

$$f(x,y) = r(x, y+y_0) + t(x-x_i, y-y_i-y_0) + n(x, y-y_0) \tag{1}$$

The input joint image is introduced into SLM1 as shown in Fig. 1. The SLMs are illuminated with the collimated light beam generated from the laser. In the first step, all pixels of SLM2 are set to 1. Lens L1 performs the Fourier transform of $f(x,y)$ and the intensity of the complex light distribution produced in the back focal plane of L1, referred to as the joint power spectrum (JPS), is typically detected by a square-law detector such as a CCD array, given b

$$\begin{aligned} |F(u,v)|^2 = {} & |R(u,v)|^2 + |T(u,v)|^2 + |N(u,v)|^2 \\ & + 2|T(u,v)||R(u,v)|\cos[\phi_t(u,v) - \phi_r(u,v) - ux_i - vy_i - 2vy_0] \\ & + 2|R(u,v)||N(u,v)|\cos[\phi_r(u,v) - \phi_n(u,v) + 2vy_0)] \\ & + 2|T(u,v)||N(u,v)|\cos[\phi_t(u,v) - \phi_n(u,v) - ux_i - vy_i] \end{aligned} \tag{2}$$

where $|R(u,v)|$, $|T(u,v)|$ and $|N(u,v)|$ are the amplitudes, and $\phi_r u,v)$, $\phi_t(u,v)$ and $\phi_n(u,v)$ are the phases of the Fourier transforms of $r(x,y)$, $t(x,y)$ and $n(x,y)$ respectively; u and v are mutually independent frequency domain variables. Using the Fourier plane image subtraction technique [11], the zero order term as well as the cross -correlation terms between

similar target or non-target objects that may be present in the input scene can be eliminated. If the input-scene-only power spectrum $|T(u,v)|^2$ and reference-image-only power spectrum $|R(u,v)|^2$ are subtracted from the JPS of Eq. (2), the modified JPS becomes,

$$\begin{aligned} P(u,v) &= |F(u,v)|^2 - |T(u,v)|^2 - |R(u,v)|^2 \\ &= 2|T(u,v)||R(u,v)|\cos[\phi_t(u,v) - \phi_r(u,v) - ux_i - vy_i - 2vy_0] \\ &\quad + 2|R(u,v)||N(u,v)|\cos[\phi_r(u,v) - \phi_n(u,v) - 2vy_0] \end{aligned} \tag{3}$$

The fringe-adjusted filter, used in a fringe-adjusted JTC [10], is defined as

$$H(u,v) = C(u,v)\left[D(u,v) + |R(u,v)|^2\right]^{-1} \tag{4}$$

where $C(u,v)$ and $D(u,v)$ are either constants or functions of u and v. When $C(u,v)$ is properly selected, one can avoid having an optical gain greater than unity. With very small values of $D(u,v)$, the pole problem otherwise associated with an inverse filter is overcome. If $C(u,v)$ = 1 and $|R(u,v)|^2 >> D(u,v)$, then Eq. (4) ca be approximated as $H(u,v) \approx |R(u,v)|^{-2}$. The modified fringe-adjusted JPS is obtained by multiplying the modified JPS of Eq. (3) by th fringe-adjusted filter function $H(u,v)$ and is thus given by

$$O(u,v) = H(u,v) \times P(u,v) \approx |R(u,v)|^{-2} \times P(u,v) \tag{5}$$

This multiplication is achieved by displaying the JPS and the FAF in SLM1 and SLM2, respectively, and then illuminating these SLMs with the c ollimated light beam. Lens L1 performs a n inverse Fourier transform ofthe fringe-adjusted JPS to yield the correlation output which consists of a pair of delta -function-like correlation peaks corresponding to the input scene target located at $(\pm x_i, \pm y_i \pm 2y_0)$, and negligible cross-correlation output between the reference and non-target objects as well as the noise term. The location of the target in the unknown input scene is determined by examining the location of the maximum correlation peak in the output plane.

3. SDF MODEL UPDATE

In general, FLIR images recorded using existing infrared cameras generate lo -resolution noisy images, and target signatures may appear very different from one frame to another. It is necessary to identify the target regardless of the distortions, which could possibly occur and to reject non-target objects with low false alarm rate while ensuring a high probability of detection. To accommodate the problem of target signature variation due to in-plane/out-of-plane rotation and scale variations, partial occlusion, noise or bad frames, an enhanced version of the SDF [12,13] concept is used. Conventional SDF formulation uses a training set of images, which characterizes the expected geometric distortions. The training set images are used to reduce the filter's sensitivity to object distortions. An enhanced SDF formulation i employed in this paper which preserves the information available from the preceding target reference images to reduce the effects of noise, hazard associated with bad frames, as well as other artifacts.

A composite SDF reference image can be generated by summing the multiplication of reference images of the target used in the previous frames with a set of weighted coefficients as shown in the following equation

$$h_n(x,y) = a_0 s_0(x,y) + a_1 s_1(x,y) + a_2 s_2(x,y) \cdots a_{n-1} s_{n-1}(x,y), \tag{6}$$

where $s_i(x,y)$ is the ith reference image with $s_0(x,y)$ representing the first reference image, and a_0, a_1, ...,a_n are a set o weighted coefficients. The composite SDF of Eq. (6) is designed in such a way that if all consecutive reference images are identical, i.e., $s = s_0 = s_1 \ldots = s_n$, then the corresponding SDF composite image would be the reference image, i.e., $h_n = s$. Under the aforementioned condition, the summation of all coefficients becomes equal to 1. To eliminate the need fo preserving all of the preceding reference images in the formulation of the SDF shown in Eq. (6), a recursive form of SDF is proposed herein, defined as

$$h_0(x,y) = s_0(x,y)$$
$$h_i(x,y) = \frac{1}{(k+1)}\left[k \times h_{i-1} + s_i(x,y)\right], \qquad i \le n \tag{7}$$

where k is an arbitrary constant. Rewriting Eq. (7) in the form of a series of summation yields

$$h_n(x,y) = \frac{k^{n-1}}{(k+1)^{n-1}} s_0(x,y) + \sum_{i=1}^{n-1} \frac{k^{n-i-1}}{(k+1)^{n-i}} s_i(x,y). \tag{8}$$

Comparing Eqs. (6) and (8), we get

$$a_i = \frac{k^{n-i-1}}{(k+1)^{n-i}}, \qquad 0 < i < n \tag{9}$$
$$= \frac{k^{n-1}}{(k+1)^{n-1}}, \qquad i = 0$$

From Eq. (9), it is evident that

$$\frac{k^{n-1}}{(k+1)^{n-1}} + \sum_{i=1}^{n-1} \frac{k^{n-i-1}}{(k+1)^{n-i}} = 1 \tag{10}$$

The larger the value of k, the more information from prior reference images is incorporated in the SDF formulation. This makes the tracking algorithm less vulnerable to the effects of bad frames. In addition, it becomes less susceptible to the changes in target appearance with respect to various 3D distortions such as rotation and scale variations. For example, if k is chosen to be equal to 2, the SDF composite image generated at frame #5 of a given image sequence will be

$$h_5(x,y) = \frac{16}{81} s_0(x,y) + \frac{8}{81} s_1(x,y) + \frac{4}{27} s_2(x,y) + \frac{2}{9} s_3(x,y) + \frac{1}{3} s_4(x,y) \tag{11}$$

If the target at frame #4 is totally corrupted by noise and the tracking algorithm misses the target at frame #4, then according to the SDF reference image of Eq. (11), 33% of the target signal will be lost. With the remaining 67% of the target reference signal, we have 67% chance of tracking the right target at frame #5. Similarly, if the target in frame #3 i also corrupted by noise, we have 44% chance of tracking the right target at frame #5.

Figure 2 illustrates an example of SDF assistance in the remedy of mistracking of a truck caused by a bad frame. Images in the left column are the input FLIR images where the tracked target is shown within a black box. Images in the middle column are the segmented reference images of the corresponding frame, while images in the right column correspond to the associated SDF composite images. In this FLIR sequence, frame #4 is totally corrupted by noise which results in the mistracking of the target. Conventional tracking methods [14-16], which utilizes only the target information from a single preceding frame, will have serious difficulty to track the target in subsequent frames without discarding the bad frame. However, the proposed tracking method using SDF based target model update effectively tracks the target following the appearance of the bad frame.

4. FJTC BASED TARGET TRACKING

Target tracking is more challenging than target detection because tracking involves the detection in consecutive frames while accommodating the effects of various 3D distortions, occlusion, noise as well as other artifacts. In general, a tracking algorithm identifies and tags the right target in a sequence of frames regardless of the target's signature variations. The proposed FJTC-based target tracking algorithm utilizes *a priori* information of the target from the previous frames such as the location of the target and the surrounding environment. The proposed tracking technique

uses multiple epochs of the FJTC tracking algorithm to zoom on a nd track the target while ensuring that the tracked target is the desired target.

In FLIR imagery, targets may consist of just a few pixels, which make the recognition of the target in the input scene challenging. Conventional way of solving this problem [14-16] is to divide the input frame into several smaller subimages and apply the detection algorithm to individual segmented subimages in order to detect the target. A multi-layered subimaging technique proposed in Reference 14 is depicted in Fig. 3(a), and requires that the target image size to be within 25% of the subimage size. It is also proposed to reduce the number of correlation operations by only correlating the subimages inside the radius of interest. Subimaging based segmentation approach [14] is inefficient because there are cases where none of the subimages may cover the entire image of the desired target as shown in Fig. 3(b). In Reference 17, a high-resolution image reconstruction technique is proposed by utilizing a number of shifted low-resolution subframes. By utilizing the reverse procedure, a frame available from a sequence is divided into a number of equal sized subframes. By shifting the grid in horizontal, vertical and diagonal directions, the entire target image can be captured in one of these subframes as shown in Figs. 3(c) through 3(f), respectively [15,16]. The size of each subframe is set equal to or greater than the size of the reference image, which is selected from a database or from a ground truth file. The entire input frame i s divided into equal sized subframes using the aforementioned procedure , and the reference subframe is used for searching the unknown target in the entire frame where the subframe is a 25x25 -pixel window. This method requires many iterations of correlation processing which are redundant because the target resides in only one of the segmented subframes. Besides, in high cluttered input scenes, it is difficult to select the accurate subframe, which contains the actual target.

In this paper, we propose an effcient way of utilizing the FJTC algorithm for tracking the target. In this method, the tracking of a target in one frame is divided into two or more processes. The FJTC based tracking is applied multiple times starting with the entire input scene as the input image with a large reference image obtained by segmentation from the preceding frame. Then the input image size and reference image size are gradually reduced until the reference image size becomes equal to the target size. In this way, unnecessary FJ TC processing is eliminated. This technique also effectively removes camera ego-motion. Multiple epoch processing not only compensates ego-motion, but also ensures efficient target tracking under complicated scenarios such as multiple independently moving objects, distortions due to skew, scale and rotation variations, as well as other artifacts.

Figure 4 shows the block diagram of the proposed FJTC algorithm for tracking one target in a sequence. The proposed technique takes two consecutive frames from the image sequence and the location of the target in the previous frame i.e., reference frame as inputs and yields the location of the target in the subsequent frame as the output. The process is then repeated with the reference frame substituted by the preceding frame and the input frame substituted by the subsequent frame in the sequence.

In the first epoch of the process, the size of the input scene is set equal to the size of input frame while the reference image size is set equal to the reference image s ize or to a size larger than the target size. If the reference image size is very small, it may result in low peak-to-side lobe ratio (PSR) in the FJTC correlation output and may not be able to compensate high ego motion of the input image. Large referenc e image is favored but requires more epochs of the FJTC process. The reference image size and the input image size are reduced in every epoch until the reference imag size reaches the target size. The total number of epochs or iterations of the FJTC proce ss (m) depen on the tracking application, especially the processing speed. If the processing speed for running one epoch is faster than the output required by the application, additional epochs are preferred because the higher the number of epochs, the better the performance of the tracking algorithm.

The image preprocessing employed in this research work includes edge enhancement, normalization, and segmentation of the reference and input images.

5. SIMULATION RESULTS

A detailed simulation program was developed using the MATLAB software package for evaluating the performance of the proposed algorithm. Real world FLIR image sequences were used for testing the performance of the proposed algorithm. The FLIR images used in the simulation involves hundreds of 128x128 -pixel gray-level images in each

sequence. Figure 5 illustrates FJTC based tracking results for one frame of a FLIR sequence. In this example frame, a very small and low contrast tank is stationed at the junction of a road when the images are taken from an infrared camera installed on a moving platform. Figure 5(a) represents the input joint image of the first tracking process containing the entire preprocessed input image and the reference image segmented from the previous frame. Figure 5(b) depicts the correlation output produced by FJTC corresponding to the input joint image of Fig. 5(a). Figure 5(c) illustrates the tracking result of the first iteration process indicating the best-fitted location of the reference image in the input scene. The outlined box i n Fig. 5(c) indicates the position and size of the input image for the next FJTC iteration process. Figure 5(d) represents the input joint image of the final tracking process of the frame while Figure 5(e) shows the corresponding correlation output. Figur 5(f) depicts the tracking result obtained indicating the final position of the target.

Figure 6 demonstrated the tracking results of the FJTC tracking algorithm for ive challenging scenarios. Th FLIR sequence in Fig. 6(a) and Fig. 6(b) illustrate the zoom in of two cold stationary tanks. The FLIR sequence in Fig. 6(c) illustrates the tracking of a tank in a low signal-to-noise ratio sequence. The FLIR sequence in Fig. 6(d) illustrates the tracking of a tank under high camera ego-motion. And the FLIR sequence i Fig. 6(e) illustrates the tracking of an independent moving object. The targets tracked by the FJTC algorithm are shown by the rectangular black box. The target always remain inside the black box for the entire sequence indicating the effectiveness of the proposed tracking algorithm.

Nevertheless, there are a few practical situations where the tracking algorithm may not function with 100% accuracy. Examples of such situations include overlapping of the target by other none target objects, and rapid temperature changes due to firing of weapons and explosions. One of the examples of overlapping target scenario is the tracking of tank behind civilian vehicle. The modified SDF model employed in this paper has the ability to mend some of those above-mentioned situations since the updated model of the target contains the feature information of the target from the previous frames. Therefore, if the problem does not propagate for too many frames and the target does not change his general appearance, continuous tracking of the target using the FJTC algorithm can still be achieved.

6. CONCLUSION

Real time recognition and tracking of targets in FLIR image sequences is a challenging problem due to low contrast, various 3D distortions, occlusion, overlapping as well as other artifacts. To alleviate these problems, a new real-time invariant target tracking algorithm is proposed using a modified SDF based FJTC algorithm. This technique can track small objects comprising of only a few pixels while compensating the cam era ego-motion. Simulation results using real life FLIR imagery verify the effectiveness of the proposed technique. Future work in this area should include data fusion of this technique with other ATR methods to solve the problem of partial object occlusi n, which is currently under investigation.

ACKNOWLEDGEMENT

This research was supported by two grants from the Army Research Office (ARO Grant Number: 43004-CI, and 43402-CI-DPS). The authors wish to thank Richard Sims and William Sander for supplying the FLIR image sequences and fo many rewarding discussions

REFERENCES

1. H. Shekarforoush and R. Chellappa, "A multi-fractal formalism for stabilization, object detection and tracking in FLIR sequences," in *IEEE Int. Conf. on Image Processing*, vol. 3, pp. 78-81, 2000
2. J. Y. Chen and I. S. Reed, "A detection algorithm for optical targets in clutter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 23, no. 1, pp. 46-59, 1987.
3. A. Strehl and J. K. Aggarwal, "Detecting moving objects in air -borne forward looking infra-red sequences," *Machine Vision Applications Journal*, vol. 11, pp. 267-276, 2000.
4. U. Braga-Neto and J. Goutsias, "Automatic target detection and tracking in forward-looking infrared image sequences using morphological connected operators," in *33rd Conference of Information Sciences and Systems*, vol. 1, pp. 173-178, March 1999.

5. D. Davies, P. Palmer and M. Mirmehdi, "Detection and tracking of very small low contrast objects," *Proceedings of the 9th British Machine Vision Conference*, pp. 599-608, BMVA Press, 1998.
6. A. Yilmaz, K. H. Shafique, M. Shah, "Target tracking in airborne forward looking infrared imagery," *Image and Vision Computing Journal*, vol. 21, pp. 623-635, 200 .
7. D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," *IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 2, pp. 142-149, 2000.
8. B. Javidi and C. Kuo, "Joint transform image correlation using a binary spatial light modulator at the Fourier plane," *Appl. Opt.*, vol. 27, pp. 663-665, 1998.
9. D. Feng, H. Zhao, and S. Xia, "Amplitude-modulated JTC for improving correlation discrimination," *Opt. Commun.*, vol. 86, pp. 260-264, 1991.
10. M. S. Alam and M. A. Karim, "Fringe-adjusted joint transform correlation", *App. Opt.*, vol. 32, no. 23, pp. 4344-4350, 1993.
11. M. S. Alam and M. A. Karim, "Multiple target detection using a modified fringe-adjusted joint transform correlator", *Opt. Eng.*, vol. 33, pp. 1610-1617, 1994.
12. M. S. Alam, X. W. Chen, and M. A. Karim, "Distortion-invariant fringe-adjusted joint transform correlation," *Appl. Opt.* vol. 36, pp. 7422-7427, 1997.
13. D. Casasent and W. T. Chang, "Correlation synthetic discriminant functions," *Appl. Opt.* vol. 25, pp. 2343-2350, 1986.
14. K. Gudmundsson, and A. Awwal, "Sub-imaging technique to improve phase only filter search capability", *Appl. Opt*, vol. 42, pp. 4709-4717, 2003.
15. Y. Wu and T. S. Huang, "Nonstationary color tracking for vision-based human-computer interaction," *IEEE Tarns. Neural Networks*, vol. 13, pp. 948-960, 200 .
16. A. K. Rastogi, B. N. Chatterji, and A. K. Ray, "Design of real-time tracking system for fast moving objects," *IETE Journal of Research*, vol. 43, pp. 359-369, 199 .
17. M. S. Alam, J. G. Bognar, R. C. Hardie and B. J. Yasuda, "Infrared image registration and high resolution reconstruction using multiple translationally shifted aliased video frames," *IEEE Transactions on Instrumentation and Measurement*, vol. 49, pp. 915-923, 2000.
18. M. S. Alam et al., "Fringe-adjusted JTC based target detection and tracking using subframes from a video sequence," to appear, *Proceedings of the SPIE*, vol. 5201, SPIE, 2003.

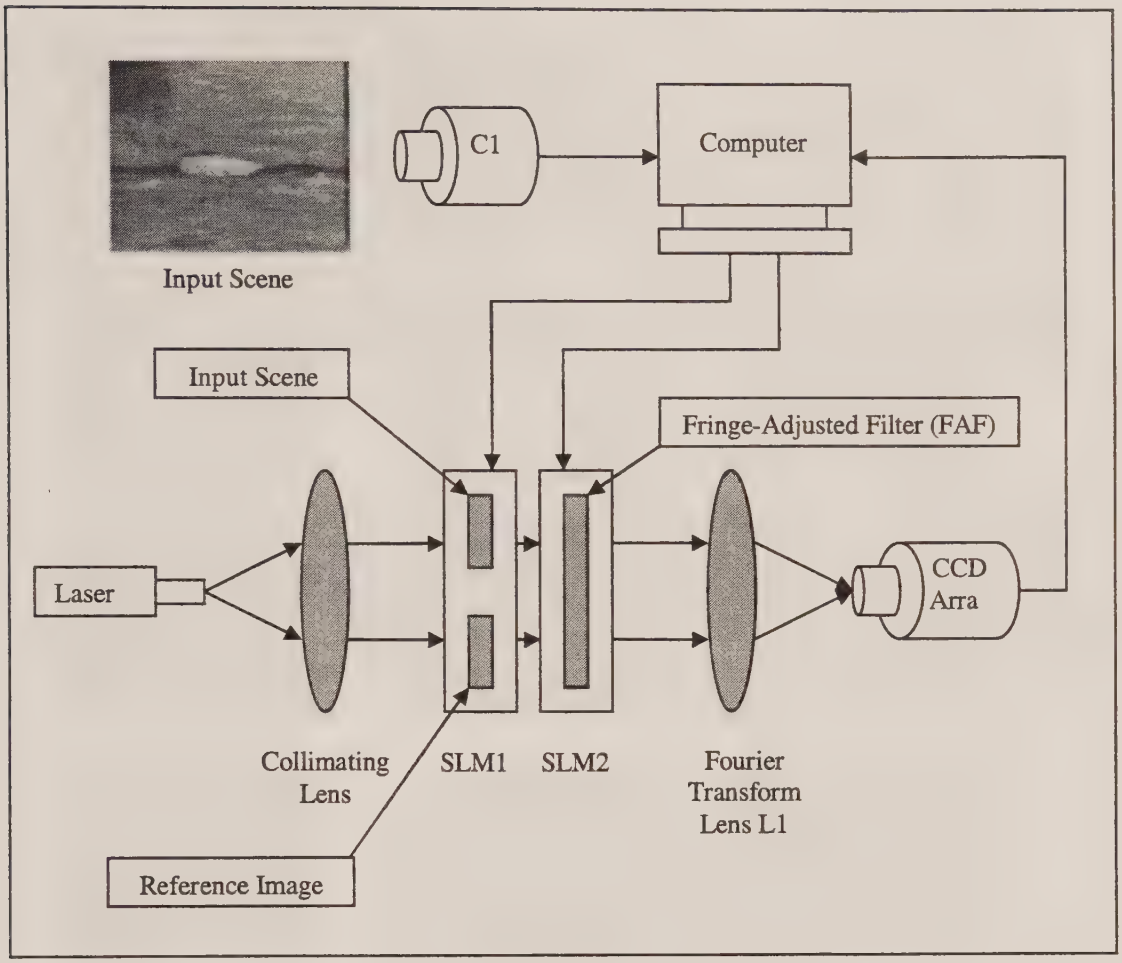


Fig. 1 Real-time fringe-adjusted JTC architecture.

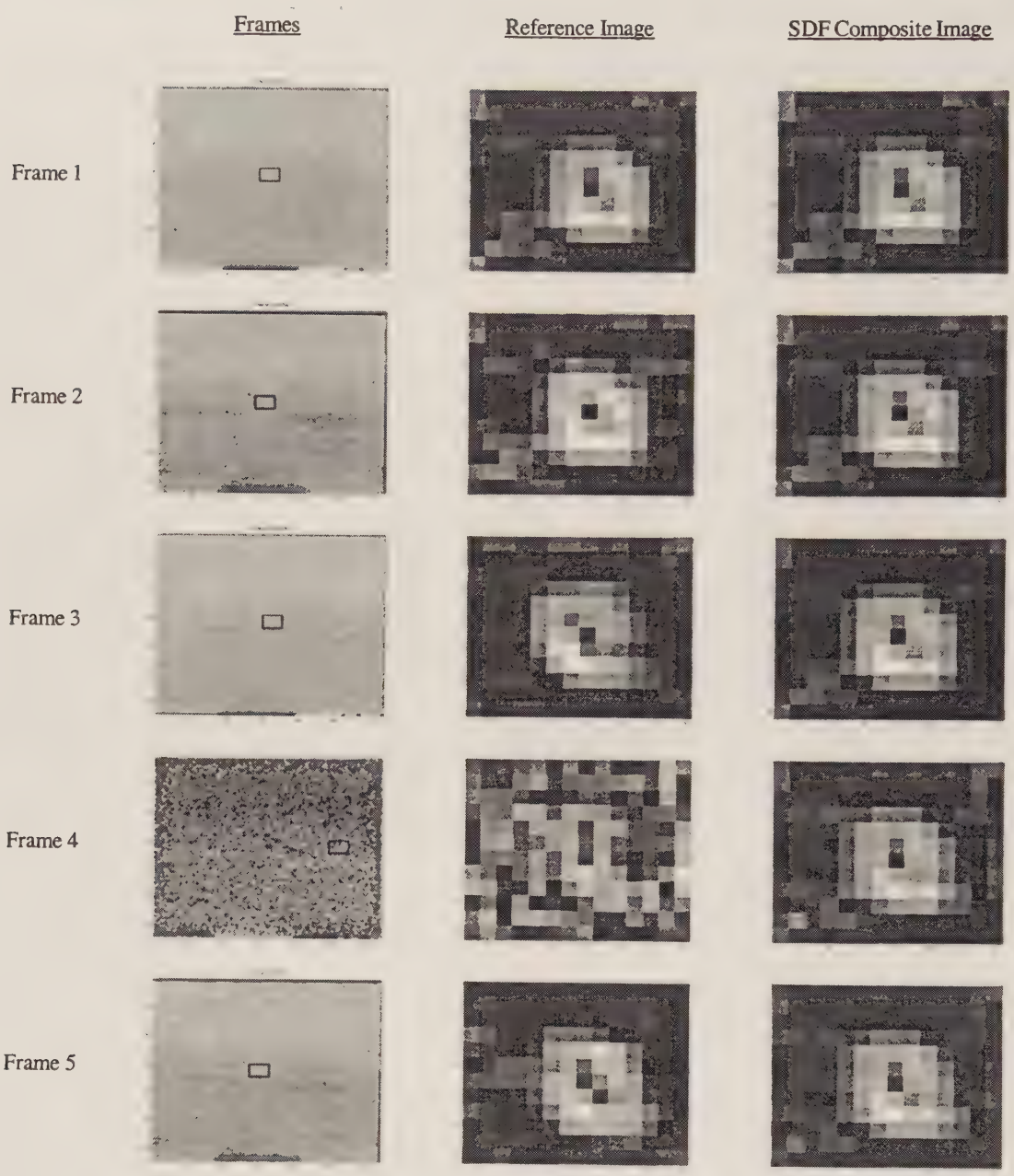
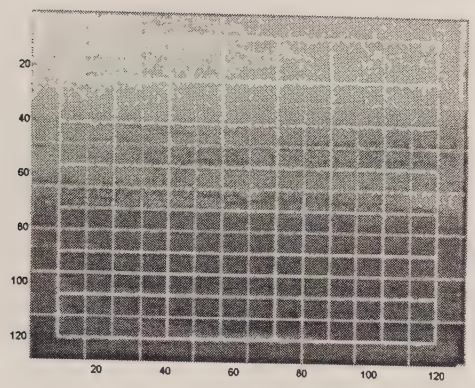
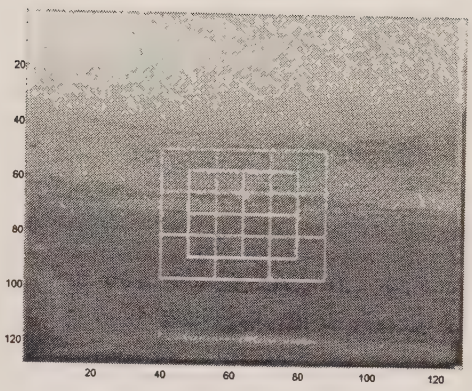


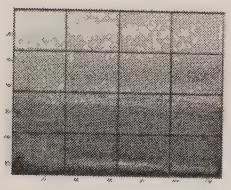
Fig. 2 Target modal update with Synthetic Discriminant Function.



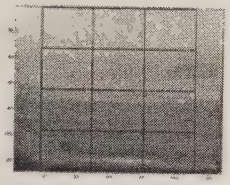
(a)



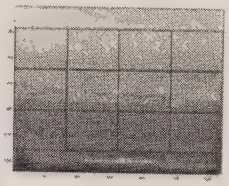
(b)



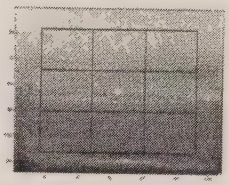
(c)



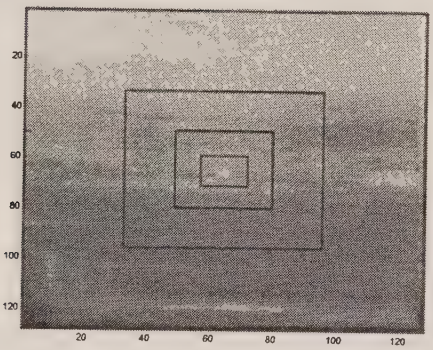
(d)



(e)



(f)



(g)

Fig. 3(a) Two-layer sub-image grid [14].

Fig. 3(b) Two-layer sub-image grid with radius of interest module.

Fig. 3(c) to (f) Forty-nine (32x32) tracking regions are divided for a single (128x128) frame by conventional grid segmentation.

Fig. 3(g) Four tracking regions are divided for a single (128x128) frame by zoom segmentation.

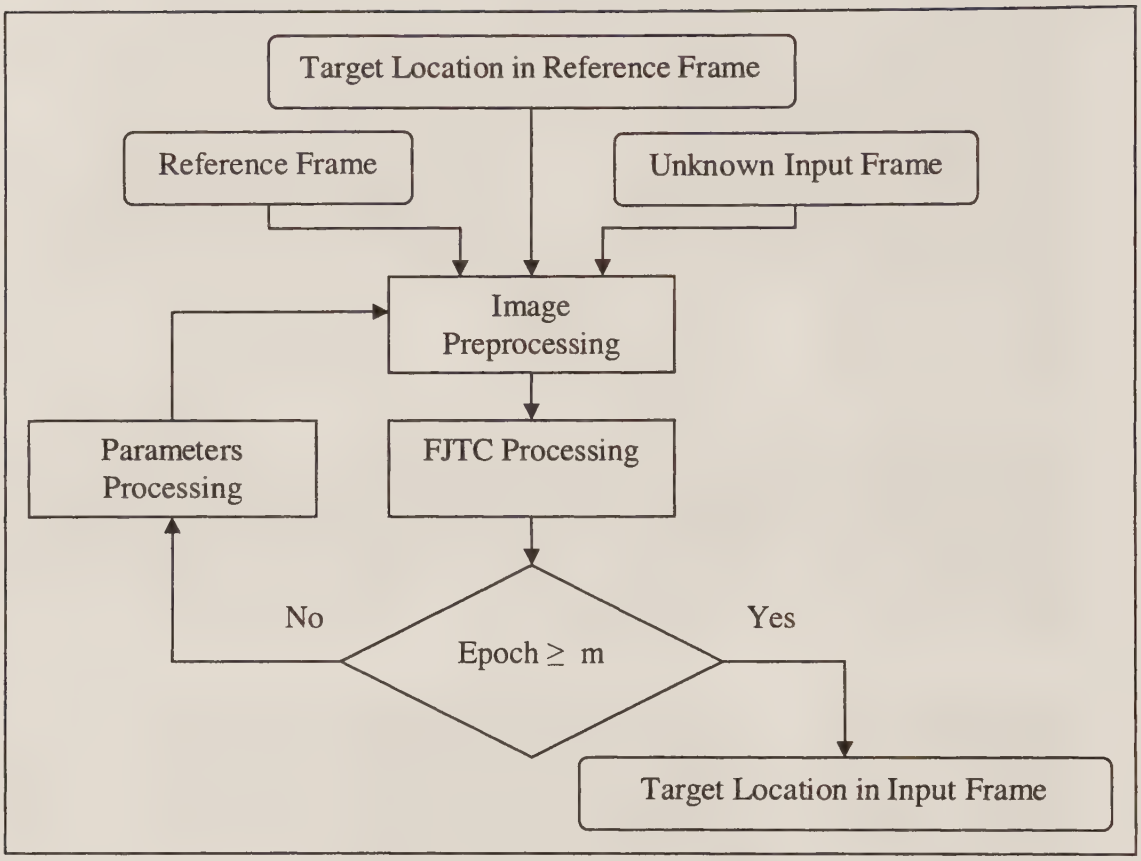
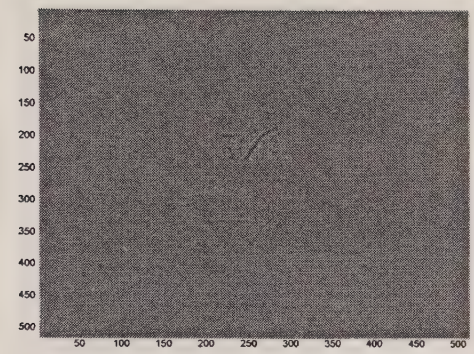
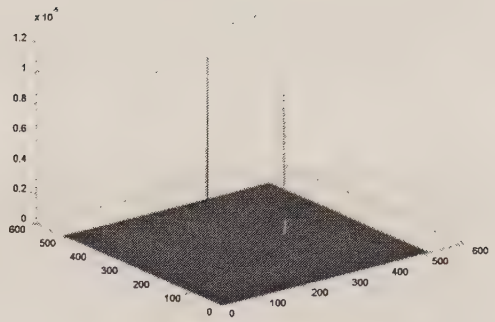


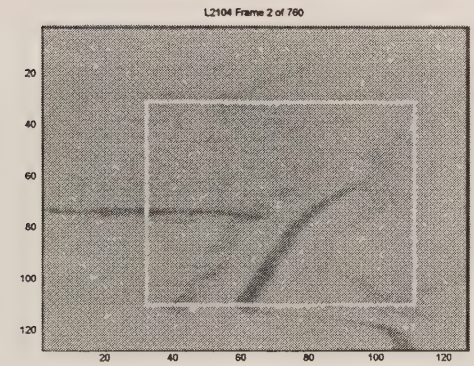
Fig. 4 Real-time Object tracking using FJTC algorithm.



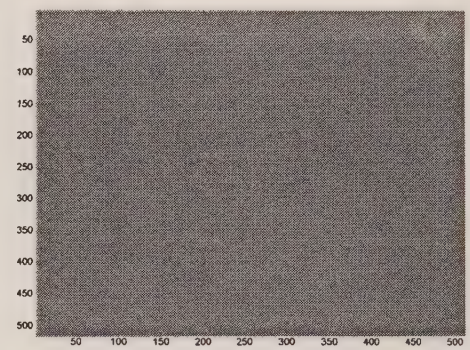
(a)



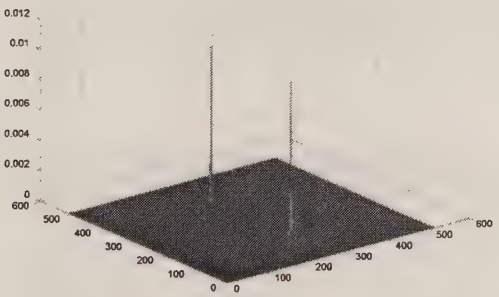
(b)



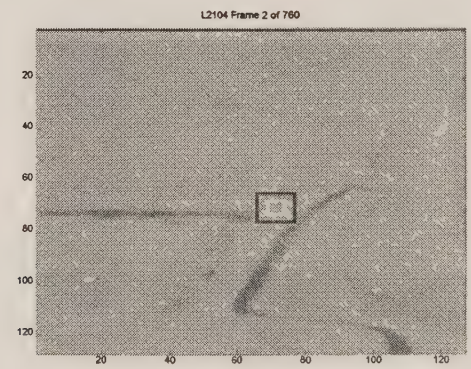
(c)



(d)

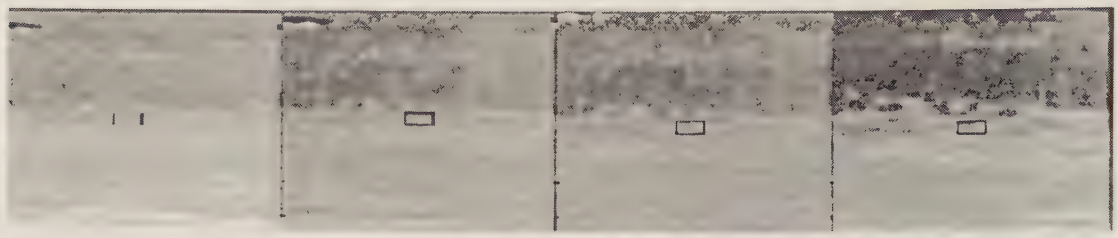


(e)

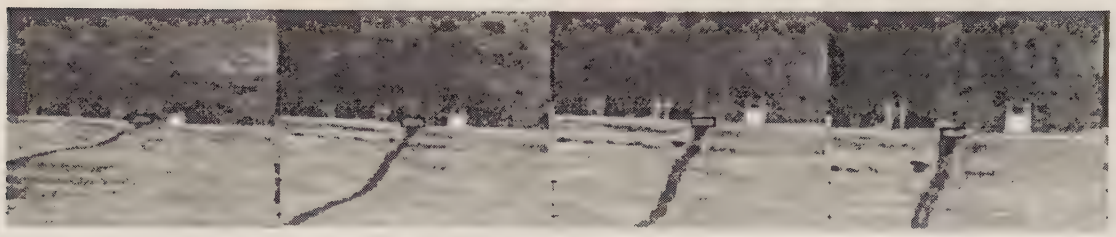


( )

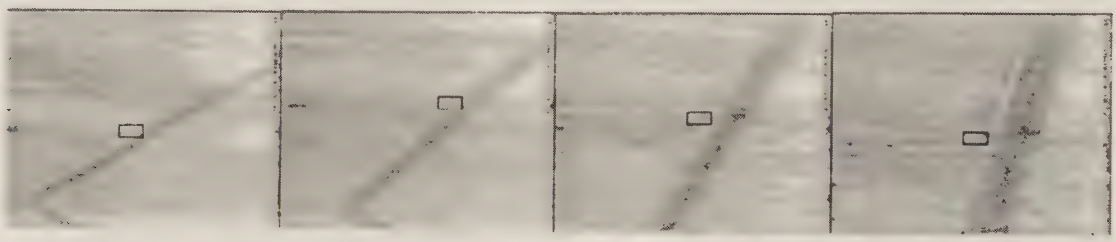
Fig. 5 FJTC based tracking of an almost invisible tank in one frame of a sequence: (a) Input joint image corresponding to the first frame of the sequence, (b) Correlation output corresponding to Fig. 1(a), (c) Target tracking using the result of Fig. 1(b), (d) Input joint image corresponding to the last frame of the sequence, (e) Correlation output corr esponding to Fig. 1(d), and (f) Target tracking using the result of Fig. 1(b) enclosed in the rectangular white box as shown in Fig. 5(f). The black box shown in Fig. 5(f) is the input image of the last process.



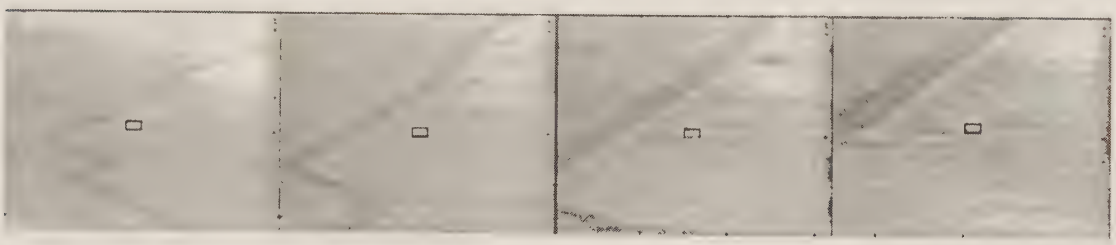
(a)



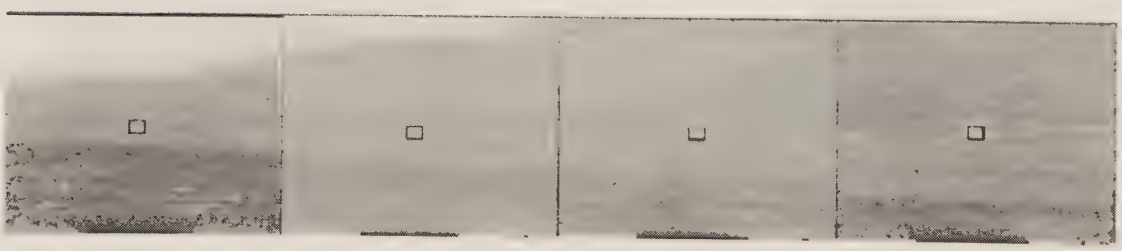
(b)



(c)



(d)



(e)

Fig. 6 Simulations results obtained using th e proposed modified SDF based FJTC tracking algorithm for four long wave and one medium wave FLIR sequences: (a) Frames #1, #50, #100 and #150 of the first sequence (L1803), (b) Frames #1, #100, #150, and #200 of the second sequence (L1902), (c) Frames #1, #100, #200 and #250 of the third sequence (L2008), (d) Frames #1, #200, #300, and #400 of the forth sequence (L2018), and (e) Frames #1, #100, #200, and #300 of the last sequence (M1413).

Performance assessment of unconstrained Hybrid Optical Neural Network (U-HONN) filter for object recognition tasks in clutter

Ioannis Kypraios, Rupert C. D. Young*, Chris R. Chatwin

Laser and Photonics Systems Research Group
Department of Engineering and Design
University of Sussex
Brighton BN1 9QT

ABSTRACT

Previously we have described a hybrid optical neural network (HONN) filter. The filter is synthesised employing an artificial neural network technique that generates a non-linear interpolation of the intermediate train set poses of the training-set objects but maintains linear shift-invariance which allows potential implementation within a linear optical correlator type architecture. In this paper, we remove the constraints imposed on the filter's output correlation peak height from the constraint matrix of the synthetic discriminant function used to create the composite filter. We examine the U-HONN filter's detectability, peak sharpness, within-class distortion range, discrimination ability between an in-class and out-of-class object and the filter's tolerance to clutter. We assess the behaviour of the U-HONN filter in an open area surveillance application. The filter demonstrates good object detection abilities within cluttered scenes, keeping good quality correlation peak sharpness and detectability throughout all the sets of tests. Thus the U-HONN filter is able to detect and accurately classify the in-class object within different background scenes at intermediate angles to the train-set poses.

Keywords: artificial neural network, non-linear superposition, synthetic discriminant function, shift-invariance, correlation filters, optical correlator, pattern recognition

* E-mail: R.C.D.Young@sussex.ac.uk; Tel: +44(0)1273 678908; Fax: +44(0)1273 690814

Optical Pattern Recognition XV, edited by David P. Casasent, Tien-Hsin Chao, Proc. of SPIE Vol. 5437 (SPIE, Bellingham, WA, 2004) · 0277-786X/04/$15 · doi: 10.1117/12.542058

1. INTRODUCTION

Mahalanobis et al.[1] in their work on the unconstrained filters[2-4] noticed that SDF-type filters[5-7], which produce broader correlation peaks but offer better distortion tolerance, were not explicitly optimised to provide good quality discrimination ability between classes. The synthetic discriminant function filter sets hard constraints[8] on the correlation peak height values of the training images. The filter has set constraints on the correlation peak heights at the origin. In effect, all the training set images, when correlated with the SDF filter, are set to produce certain prespecified peak values; however, there is no information provided for the test image correlation peak height values. In contrast, in the unconstrained correlation filter synthesis there are no hard constraints on the correlation peak heights. The assumption made is that by removing the hard constraints, the number of possible solutions the filter can draw on increases by allowing the correlation peak height values to move freely to any value, so improving its performance.

Section 2 summarises the U-HONN filter's implementation. Section 3 assesses the filter's performance[9-11] based on the peak sharpness and detectability metric, the distortion range metric and the filter's discrimination ability. Emphasis is given to the performance assessment of the unconstrained hybrid optical neural network filter for recognising the object in clutter. Section 4 concludes.

2. THE UNCONSTRAINED HONN (U-HONN) FILTER IMPLEMENTATION

In an effort to keep consistency between the different mathematical symbology of artificial neural networks[12, 13] and optical correlators we will unify their representation. We denote the variable names and functions by non-Italic letters, the names of the vectors by Italic lower case letters and the matrices by Italic upper case. The frequency domain vectors, matrices, variable names and functions are represented by bold letters and the space domain vectors, matrices, variables and functions by plain letters.

If we do not set hard constraints on the correlation peak heights generated by the hybrid optical neural network filter, and add the transformed images $S_{i=1\cdots N}$, without any hard constraints weights at the origin, we synthesise the *unconstrained* hybrid optical neural network (U-HONN) filter[14, 15]:

$$\mathrm{U-HONN} = \sum_{i=1\cdots N}^{N} S_i(m,n) \tag{1}$$

where

$$S_{i=1\cdots N} = \Gamma_{i=1\cdots N} \cdot X_{i=1\cdots N}(m,n) \tag{2}$$

$$\Gamma_{i=1\cdots N} = W^{x_i} \cdot L^{x_i}$$

$$= \begin{bmatrix} w_{11}^{x_i} & w_{12}^{x_i} & \cdots & w_{1n-1}^{x_i} & w_{1n}^{x_i} \\ w_{21}^{x_i} & w_{22}^{x_i} & \cdots & w_{2n-1}^{x_i} & w_{2n}^{x_i} \\ \vdots & & & & \\ w_{m1}^{x_i} & w_{m2}^{x_i} & \cdots & w_{mn-1}^{x_i} & w_{mn}^{x_i} \end{bmatrix} \cdot \begin{bmatrix} l_{11}^{x_i} & \cdots & l_{1q}^{x_i} \\ l_{21}^{x_i} & \cdots & l_{2q}^{x_i} \\ \vdots & & \\ l_{n1}^{x_i} & \cdots & l_{nq}^{x_i} \end{bmatrix}$$

$$= \begin{bmatrix} w_{11}^{x_i} l_{11}^{x_i} + w_{12}^{x_i} l_{21}^{x_i} + \cdots + w_{1n}^{x_i} l_{n1}^{x_i} & \cdots & w_{11}^{x_i} l_{1q}^{x_i} + w_{12}^{x_i} l_{2q}^{x_i} + \cdots + w_{1n}^{x_i} l_{nq}^{x_i} \\ w_{21}^{x_i} l_{11}^{x_i} + w_{22}^{x_i} l_{21}^{x_i} + \cdots + w_{2n}^{x_i} l_{n1}^{x_i} & \cdots & w_{21}^{x_i} l_{1q}^{x_i} + w_{22}^{x_i} l_{2q}^{x_i} + \cdots + w_{2n}^{x_i} l_{nq}^{x_i} \\ \vdots & & \\ w_{m1}^{x_i} l_{11}^{x_i} + w_{m2}^{x_i} l_{21}^{x_i} + \cdots + w_{mn}^{x_i} l_{n1}^{x_i} & \cdots & w_{m1}^{x_i} l_{1q}^{x_i} + w_{m2}^{x_i} l_{2q}^{x_i} + \cdots + w_{mn}^{x_x} l_{nq}^{x_x} \end{bmatrix} \tag{3}$$

where W^{x_i} and L^{x_i} are the matrices of the input and layer weights[13]. $w_{mn}^{x_x}$ are the input and layer weights from the input neuron of the input vector element at row m and column n to the associated hidden layer for the training image $x_{i=1\ \ldots\ N}(m,n)$. $l_{nq}^{x_i}$ are the input and layer weights from the hidden neuron of the layer vector element at row m and column n to the associated output neuron. In our case, q = 1 since the output layer has only one neuron. $S_{i=1\ldots N}(m,n)$ is the transformed image calculated from the dot product of $\Gamma_{i=1\ \ldots\ N}$ with the corresponding training image $X_{i=1\ldots N}(m,n)$. The non-linear U-HONN filter is inherently shift invariant and may be employed as an optical correlator-type filter. It may be used as a space domain function in a joint transform correlator architecture or be Fourier transformed and used as a Fourier domain filter in a 4-f Vander Lugt type optical correlator[16].

3. PERFORMANCE OF THE UNCONSTRAINED HYBRID OPTICAL NEURAL NETWORK (U-HONN) FILTER

The size of all the images us $[\,256 \times 256\,]$ and the grey-scale bitmap format is used. All the images of the training set for the artificial neural network (NNET) block of the U-HONN filter as well as for the correlator block are concatenated row by row into a vector of size $[\,1 \times (\,256 \times 256\,)\,]$ prior to input to the neural net. Normally this size of image is not possible, since to be implemented by enough input and layer weights we would require:

$$\begin{aligned} N_{iw} &= 10 \times [\,256 \times 256\,] \\ &= 10 \times 65{,}536 \\ &= 655{,}360 \end{aligned} \tag{4}$$

This would be impossibly large since for a training set of e.g. N = 10 images of individual vector size $[\,256 \times 256\,]$ since there would be more than half-a-million input weight connections[10] needed. Thus the selective weight connection architecture is employed to overcome this problem[14,15]. Also, introducing into the NNET

training session the heuristic training algorithm with momentum and an adaptive learning rate[17, 18] has speeded up the learning phase and reduced the memory size needed to complete fully the training session. The NNET typically converges for each training image in under one hundred iterations, requiring only a few tens of milliseconds CPU time on a 2.5GHz Pentium 4 based PC. Also, due to the generalization properties[12, 13] of a NNET structure, the number of the training images decreases in comparison to the typical number of images in the training set of linear correlator-type filters[8].

3.1 Peak sharpness and detectability[9-11]

We assessed the ability of the U-HONN filter to detect non-training in-class images that are oriented at the intermediate angle of view between the training images. The training set consisted of images over a distortion range $[20 \quad 80]$ degrees at increments of 20°. We tested the unconstrained hybrid optical neural network filter with the intermediate car poses over the same range at 10° increments. We added two randomly chosen intermediate car poses, at 130° and at 140°, in the training set of the filter to create a false class. The Target[14, 15] of the false class object is $T_{false} = -40$ and of the true class object is $T_{true} = +40$. The U-HONN filter has no information on the non-training, intermediate car images in the construction of its composite image. There were no constraints of the correlation peak set explicitly by the constraint matrix.

Fig. 1 plots the correlation-peak height for each input image for the unconstrained hybrid optical neural network filter. From the plot it is apparent that the U-HONN is invariant to the out-of-plane rotation, since it has produced correlation peaks for both the in-class training and non-training images around the average auto-correlation peak-height value equal to 1×10^3. Although there is a greater variation on the peak-height values of the U-HONN filter than of the constrained hybrid optical neural network filter (C-HONN), both filters demonstrated invariance to the out-of-plane rotation. Again, as for the constrained hybrid optical neural network filter, the consistency of the correlation peak values produced from the unconstrained hybrid optical neural network filter demonstrate the filter's ability to interpolate well between the intermediate car poses at 10° increments. The improvement in the U-HONN (as well as in the C-HONN) filter's interpolation ability must result from the non-linear function applied in the construction of the composite image comprising the SDF filter formed from the NNET output images. It can be shown[19] that the NNET interpolates non-linearly between the reference images included in the training set and forces all the non-reference objects to follow the activation graph:

$$o[u(i+1)] = \xi \frac{1 - k\ exp(\mu a_1 u_1)\ exp(\mu a_2 u_2)\ exp(\mu a_3 u_3) \cdots exp(\mu a_{N-1} u_{N-1})\ exp(\mu a_N u_N)}{1 + k\ exp(\mu a_1 u_1)\ exp(\mu a_2 u_2)\ exp(\mu a_3 u_3) \cdots exp(\mu a_{N-1} u_{N-1})\ exp(\mu a_N u_N)} \tag{5}$$

where o is the output of a non-linear artificial neural network (SINA network) corresponding to an input $u \in R^{+m}$, $k = \exp(-\beta r)$, $\mu\ (= \beta\lambda)$ and $\lambda = \dfrac{(r+1)}{b(i)\ v}$ are constants and $v = \|u\|$.

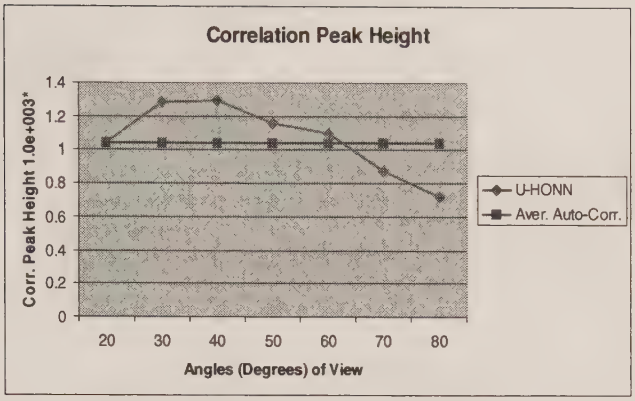


Fig. 1 The plot shows the correlation peak height, of the training set over the orientation range $[20\ 80]$ at increments of 20° for the unconstrained hybrid optical neural network filter. The test consisted of the car images over the same range, at increments of 10°. We have plotted the average auto-correlation peak-height value as a horizontal line to observe more easily the change of the test set correlation peaks. There are no constraints set on the correlation peak of the training set images.

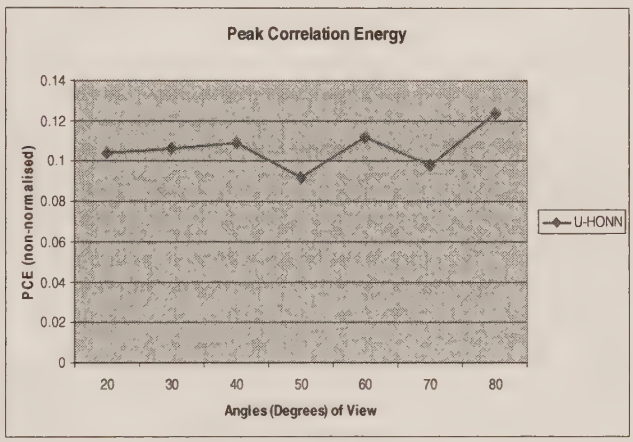


Fig. 2 The plot shows the peak correlation energy (PCE) over the orientation range $[20\ 80]$ at increments of 20° of the training set for the U-HONN filter. The test set consisted of the car images over the same range, at increments of 10°. The PCE values were not normalised.

$$\bar{x} = \frac{1}{N} f_{\kappa}(x_i) \tag{6}$$

where $\bar{x}$ is the average training set image and f( ) is the activation or transfer function[12] of node κ in the space domain. The transfer function of the hidden layers is set as the Log-Sigmoidal function, whereas the transfer function of the output layer is linear. The above equation is written in the frequency domain as:

$$\bar{x} = \frac{1}{N} f_{\kappa}(x_i) \tag{7}$$

Fig. 2 shows the non-normalised peak-to-correlation energy [9-11] (PCE) values for the unconstrained hybrid optical neural network filter. From the graph, it can be observed that the U-HONN filter produced PCE values for the intermediate non-training images close to those produced by the training car images. In effect, the filter maintains correlation peak sharpness for the in-class training images. The U-HONN filter demonstrated higher peak-to-correlation energy values than the C-HONN filter values. Thus, although the correlation-peak-height values of the C-HONN filter did not vary as much as those of the U-HONN filter due to the fixed constraints on the peaks of the C-HONN filter, allowing the filter peaks to move freely produced higher sharpness of the peaks.

From our experiments, we found that the change of the absolute distance of the classification levels ΔCl, between the true class objects and the false class objects, affected the unconstrained hybrid optical neural network filter's behaviour [20, 21] in the same way as for the C-HONN filter. When we increased the distance, the PCE values increased and, conversely, when we decreased it, the PCE values produced decreased. Similarly, as for the C-HONN filter, as we increase the value of ΔCl, the U-HONN filter increases its discrimination ability; but when we decrease ΔCl the filter increases its generalisation ability and so its discrimination ability deteriorates. The correlation peak broadens as ΔCl is decreased i.e. there is a very similar trade-off in discrimination and generalisation ability to that found in unconstrained type correlation filters. Thus increasing ΔCl leads to an increased emphasis of the high spatial frequency content of the composite images comprising the filter, which in turn leads to a more localised correlation response.

3.2 Distortion range of the filter [9-11]

Here, we evaluate the distortion range of the unconstrained hybrid optical neural network filter. The training set consisted of images for a distortion range over 0° to 90°. We used several smaller test sets, which consisted of two in-class training images at a widely separated angle within the range $\lfloor 10^\circ\ 20^\circ\ 30^\circ\ 40^\circ\ 50^\circ\ 60^\circ\ 70^\circ\ 80^\circ\ 90^\circ \rfloor$, and a third non-training in-class image lying on the bisector angle of the two in-class training images. The intermediate non-training car pose image (see Fig. 3) is $\hat{\Theta}_3 \in \lfloor 5^\circ, 45^\circ \rfloor$ i.e. $\Theta_3 = \lfloor 5^\circ\ 10^\circ\ 15^\circ\ 20^\circ\ 25^\circ\ 30^\circ\ 35^\circ\ 40^\circ\ 45^\circ \rfloor$. We added three randomly chosen training images, namely at orientations of $110^\circ, 130^\circ$ and 140°, in the training set of the unconstrained hybrid optical neural network filter (which fall inside the false class). The Targets of the true class objects and of the false class objects are $T_{true} = +40$ and $T_{false} = -40$. The filter has no information built into it on the test images of the intermediate car poses. There are no constraints set on the correlation peak values either for the in-class images or for the out-of-class images.

Fig. 4 shows the percentage of similarity of the correlation-peak height for the U-HONN filter for each input image with respect to the average auto-correlation peak-height value of the two training set images at angles Θ_1 and Θ_2. The percentage is given by:

$$\textbf{Similarity \%} = 100 - \frac{\textbf{Diff} \times 100}{\textbf{Aver Auto\_Corr Peak Height}} \tag{8}$$

where **Diff** and **Aver Auto_Corr Peak Height** are defined to be:

$$\textbf{Diff} = \textbf{Corr Peak Height} - \textbf{Aver Auto\_Corr Peak Height} \tag{9}$$

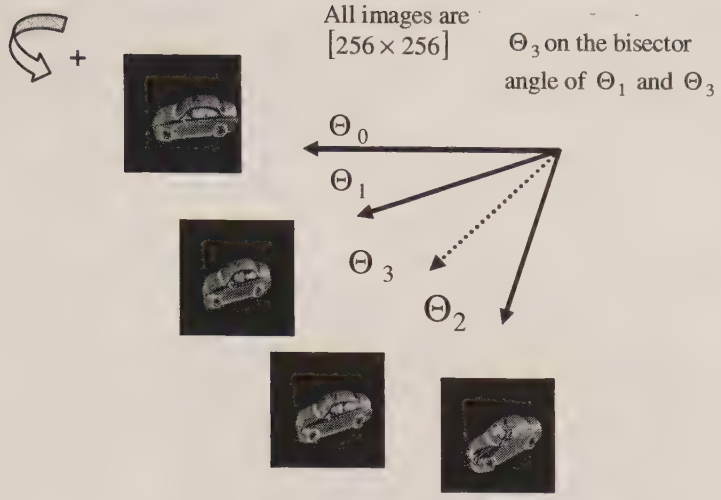


Fig. 3 The diagram shows the reference angle, Θ_0, and the two in-class training images at the angles, Θ_1 and Θ_2. The test image is at the bisector at angle Θ_3.

$$\textbf{Aver Auto \_ Corr Peak Height} = \frac{S_1 + S_2}{2} \tag{10}$$

where S_1 and S_2 are the training set images in the frequency domain at angles Θ_1 and Θ_2 respectively. It is apparent from the graph that the filter has good performance in recognising all the intermediate car poses of the test set. The correlation-peak height of almost all the in-class test set images, which are intermediate between two training images, lie within a band of greater than 90% of the average auto-correlation peak height of these two images. In effect, it is clear that the U-HONN filter demonstrates invariance to rotation over a range of $\hat{\Theta}_3 \in \left\lfloor 5^\circ, 45^\circ \right\rfloor$. Thus the U-HONN filter performs slightly better than the C-HONN filter with respect to the distortion range results, since the constrained hybrid optical neural network filter produced correlation peak-height values of the in-class test set images which are intermediate between two training images falling within a band of 85%. This is due to removing the fixed constraints on the correlation peaks. The U-HONN filter allows the correlation peak heights to move freely, in effect maximising the average auto-correlation peak-height (since some of the correlation peaks of the test images exceeded the trained images peak-height values).

3.3 Discrimination ability [9-11]

The main aim of the third group of test sets was to evaluate the ability of the U-HONN filter to discriminate between objects of different classes while retaining invariance to in-class distortions. The training set consisted of images of the Jaguar S-type for a distortion range over 20° to 70° at 10° increments. The test set consisted of one training image at 40° orientation of the Jaguar S-type and a second image of the out-of-class Police RX-7 at the same angle of view. We added two images at 130° and 140° for the false class of the filter's training set. Again, there are no constraints set on the correlation peak-height values, either of the true class images or of the false class images. The Target of the false class object is, $T_{false} = -40$, and the Target of the true class object is, $T_{true} = +40$. As before, there is no information on the test images of the in-class object or the out-of-class object provided to the unconstrained hybrid optical neural network filter.

Table 1 was drawn from the tests we performed for the U-HONN filter. The third column of the table contains the values taken for the in-class training image and the fourth column contains all the values taken for the out-of-class

training image. It can be seen from the Table 1 entries that the filter sufficient discrimination ability between the two objects, the Jaguar S-type car and the RX-7 Police patrol car. The U-HONN filter produced 12% or higher class separation. By increasing the absolute distance of the classification levels ΔCl from 80 to 200, in effect increasing the Target of the false class object to $T_{false} = -100$, and the Target of the true class object to $T_{true} = +100$, the U-HONN filter increased the class separation to 36% or higher. The C-HONN filter for $T_{false} = -40$ and

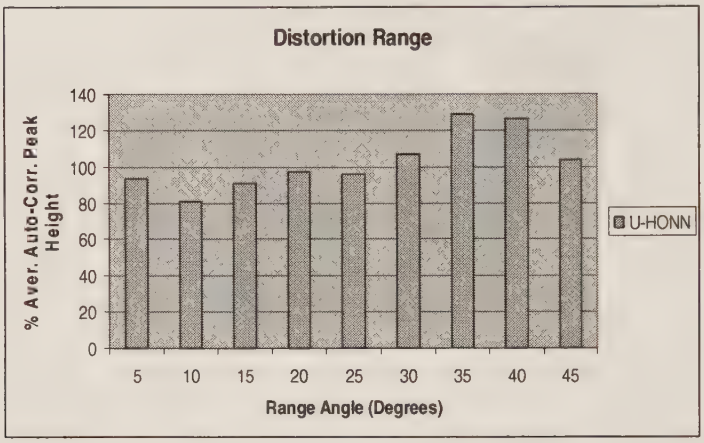


Fig. 4 The bar plot shows the similarity of the test image correlation peak-height with respect to the average auto-correlation peak height of the two training set images at angles Θ_1 and Θ_2.

$T_{true} = +40$ achieved 72% or higher class separation. Clearly, the constrained hybrid optical neural network filter produced better performance in separating the two different class objects. The unconstrained hybrid optical neural network filter has maximised the correlation peak height value of both classes of objects, minimising their class separation. In the C-HONN filter we have hardwired peak constraints for both classes. The main aim was to suppress the false class object to

Table 1 Discrimination Ability (%) of the U-HONN filter for the in-class training Jaguar S-type image at 40° and the out-of-class non-training Mazda RX-7 image at the same angle.

			Correlation Peak Height	$* 1.0 \times 10^3$
Filter	Parameter Values	Discr. Ability %	JAGUAR	MAZDA
HONN	$\Delta Cl_1 = 80$	12.75	1.6464	1.4366
HONN	$\Delta Cl_2 = 200$	36	0.6531	1.021

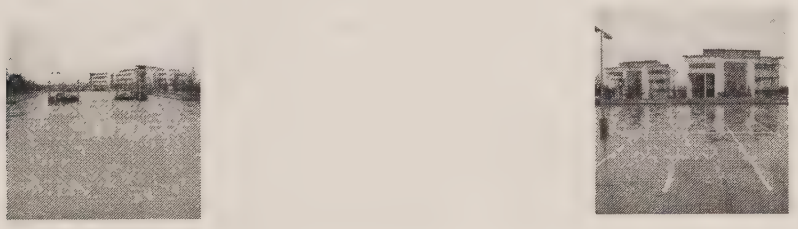
zero and fix the true class object to unity. In the U- HONN filter there are no explicit constraints set for either of the false class object or the true class object.

3.4 Clutter tolerance [9-11]

The fourth group of tests was carried out to evaluate the tolerance of the filter to background clutter in the input scene (see Fig. 5) by the insertion of training images and non-training out-of-class images into different car park

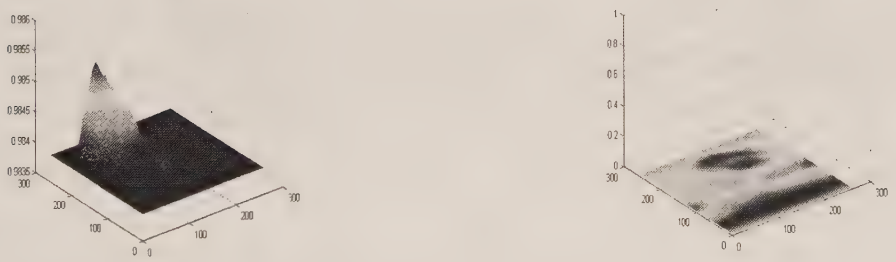
scenes. The training set consisted of images of the Jaguar S-type for a distortion range over 20° to 70° degrees at 10° increments. We added three background images of typical empty car park scenes to fall inside the false class in the training set of the unconstrained hybrid optical neural network filter. However, there were none of the false class images added in the synthesis of the final filter's composite image and there were no constraints set on the correlation peak-height values. The test set consisted of one of the training in-class images inserted in a car park scene, a randomly chosen non-training in-class image at an intermediate pose inserted in the same car park scene and one out-of-class non-training image of the RX-7 Police patrol car also inserted in the car park scene. The rest of the test set images were the same images of the Jaguar S-type and the RX-7 Police patrol car inserted in a second car park scene. The Targets of the true class objects were $T_{true} = +40$ and of the false class background images of the empty car park scenes were $T_{false\,1} = -40$. The false class image of the out-of-class non-training RX-7 Police car image was $T_{false\,2} = -40$. Fig. 6 shows the isometric correlation planes of the in-class training Jaguar S-type car image at 40° and of the out-of-class non-training RX-7 Police patrol car at the same angle of view (both resized and inserted into the first car park scene). Fig. 6 (a) shows the correlation plane of the in-class training Jaguar S-type car image at 40° and (b) shows the correlation plane of the out-of-class non-training RX-7 Police Patrol car at 40° for the U-HONN filter response. The out-of-class object's isometric plot is normalised to the maximum correlation peak-height value of the in-class object's correlation plane. Fig. 7 shows the isometric correlation planes of the in-class training view Jaguar S-type car image at 55° (a), resized and inserted in the first car park scene and, (b) resized and inserted in the second car park scene for the U-HONN filter. Both isometric plots are normalised to the maximum correlation peak height value of both of their correlation planes.

From the above results, it is apparent that the U-HONN filter was able to detect and classify correctly the Jaguar S-type car pose at 40° and suppress the background clutter scene. The filter had no previous knowledge of the



(a) (b)

Fig. 5 Two typical background scenes used in our experiments.



(a) (b)

Fig. 6 The isometric correlation planes of (a) the in-class training Jaguar S-type car image at 40° and (b) the out-of-class non-training RX-7 Police patrol car at the same angle of view (both resized and inserted in the first background scene.

Fig. 7 The isometric correlation planes of the in-class training view Jaguar S-type car image at 55° (a) resized and inserted in the first car park scene and (b) resized and inserted in the second car park scene for the U-HONN filter. The filter produced almost identical correlation planes for both background scenes.

cluttered scene images of the objects used in the test set. In effect, it was able to recognise the true-class object inserted in an unknown car park scene. The U-HONN filter was able to separate the out-of-class object from the in-class object in both background scenes, but the class separation was better in the first background scene. This is due to the higher complexity of the second car park scene in comparison to the first one. The filter's good performance is maintained with the in-class training set Jaguar car image as well as with the intermediate in-class in-training Jaguar S-type car image. It is clear from the plots of the C-HONN filter and the U-HONN filter that the constrained hybrid optical neural network filter produced sharper peaks than the unconstrained filter. However, the U-HONN filter produced higher correlation peak-height values for the test set images. Table 2 shows the correlation peak-height values of the in-class (training and non-training) and out-of-class objects in the two background scenes. By removing the constraints on the correlation peak-heights, the U-HONN filter has increased the test correlation peak-height values. In effect the filter was forced to maximise the correlation peak-height values but broaden the sidelobes of the test set images in order to be able to detect the in-class object in the cluttered scenes.

Table 2 The correlation peak-height values and the discrimination ability % of the in-class (training and non-training) and out-of-class objects in the two background scenes.

	IN-CLASS	IN-CLASS	IN-CLASS	IN-CLASS	OUT-of-CLASS
	40°	40°	55°	55°	MAZDA at 40°
BG Scenes	Corr. Peak-Height	Discrimination Ability	Corr. Peak-Height	Discrimination Ability	Corr. Peak-Height
	1.0×10^5	%	1.0×10^5	%	1.0×10^5
BG1	3.8874	90.16309101	3.8858	90.15904061	0.3824
BG2	3.8987	0.733577859	3.8885	0.473190176	3.8701

4. CONCLUSION

In this paper we have assessed the performance of the unconstrained hybrid optical neural network filter (U-HONN) filter. In the unconstrained correlation filter synthesis there are no hard constraints on the correlation peak heights. The assumption made is that by removing the hard constraints, the number of possible solutions the filter can draw on increases by allowing the correlation peak height values to move freely to any value, so improving its performance. We have assessed the performance of the U-HONN filter based on several performance metrics. The training set images of the NNET block as well as of the overall filter consist of several images of the object at different poses and at regular angular increments. Each image is passed through the NNET in gray-scale form. The

filter has produced correlation peak-height values for both the in-class training and non-training images around the average auto-correlation peak-height value. In effect it has demonstrated invariance to out-of-plane rotation. The U-HONN filter has demonstrated invariance to rotation over the range of $\hat{\Theta}_3 \in \left[5^\circ, 45^\circ \right]$. The filter has performed better than the C-HONN filter with respect to the distortion results, since the U-HONN filter has produced correlation peak-height values for almost all the in-class test set images, which are intermediate between two training images, falling within a band of greater than 90%, whereas the C-HONN filter fell within a band of greater than 85%. However, the constrained hybrid optical neural network filter has produced better class separation performance achieving 72% discrimination ability, in comparison to the U-HONN filter which has not been able to produce higher than 36% discrimination ability even when we increased the absolute distance of the classification levels. The unconstrained hybrid optical neural network filter has been able to separate the out-of-class object from the in-class object in both background scenes, but the class separation deteriorated more in the second background scene due to its higher complexity. The images of the objects in the clutter backgrounds used in the test set have not previously been presented to the filter. The filter maintained good performance with the in-class training set Jaguar car image as well as with the intermediate in-class in-training Jaguar S-type car image. Similar to the C-HONN filter, as we increase the value of ΔCl, the U-HONN filter increases its discrimination ability, but when we decrease ΔCl the filter increases its generalisation ability and so its discrimination ability deteriorates. The correlation peak broadens as ΔCl is decreased i.e. there is a very similar trade-off in discrimination and generalisation ability to that found in unconstrained type correlation filters.

ACKNOWLEDGEMENT

The principal author would like to thank everyone in the 'Laser and Photonics Systems Research Group' for contributions to the completion of this paper. The author wishes the work presented here and any related material to benefit and not harm human life: "*The time will come when the devil will make his turns with his pumpkin.* " *(Prophecy by St. Kosmas Aitolos, taken from http:// www. orthodoxheritage.org*)

REFERENCES

1. A. Mahalanobis, B. V. K. Vijaya Kumar, S. Song, S. R. F. Sims, J. F. Epperson, "Unconstrained correlation filters", Applied Optics, **Vol. 33**, 3751-3759 (1994).
2. A. Mahalanobis and B. V. K. Vijaya Kumar, "Optimality of the maximum average correlation height filter for detection of targets in noise", Optical Engineering **Vol. 36, No. 10,** 2642-2648 (1997).
3. H. Zhou and T.-H. Chao, "MACH filter synthesizing for detecting targets in cluttered environment for gray-scale optical correlator", *SPIE* Vol. 3715 394-398, Florida (1999).
4. T.-H. Chao, G. Reyes and Y. Park, "Grayscale Optical Correlator", *SPIE,* Vol. 3386 60-64 (1998).
5. D. Casasent, "Unified synthetic discriminant function computational formulation", Applied Optics, **Vol. 23, No. 10,** 1620-1627 (1984).
6. H. J. Caulfield, "Linear combinations of filters for character recognition: a unified treatment", Applied Optics, **Vol. 19, No. 23,** 3877-3878 (1980).
7. H. J. Caulfield and W. T. Maloney, "Improved discrimination in optical character recognition", Applied Optics, **Vol. 08, No. 11,** 2354-2356 (1969).
8. E. Stamos, "*Algorithms for designing filters for optical pattern recognition*", D. Phil. Thesis, University of London, (2001).
9. B. V. K. Vijaya Kumar and L. Hasserbrook, "Performance measures for correlation filters", Applied Optics **Vol. 29, No. 20** 2997-3001 (1990).
10. Ph. Refregier, "Filter design for optical pattern recognition: multicriteria optimization approach", Optics Letters, **Vol. 15, No. 15,** 854-856 (1990).
11. Ph. Refregier. "Optimal trade-off filters for noise robustness, sharpness of the correlation peak, and Horner efficiency", Optics Letters, **Vol, 16, No. 11,** 829-831 (1991).

12. M. T. Hagan, H. B. Demuth and M. H. Beale, "*Neural Network Design*", Boston, MA: PWS Publishing (1996).
13. C. G. Looney, "*Pattern recognition using neural networks: theory and algorithms for engineers and scientists*", Oxford University
Press Inc, USA, (1997)
14. I. Kypraios, R. C. D. Young, P. Birch, C. Chatwin, "A non-linear training set superposition filter derived by neural network training methods for implementation in a shift invariant optical correlator", *SPIE*, Vol.5106 84-95 (2003).
15. I. Kypraios, R. C. D. Young, P. Birch, C. Chatwin, "Object recognition within cluttered scenes employing a Hybrid Optical Neural
Network (HONN) filter", **OE Special Issue on Target Recognition-Accepted for publication** (2004)
16. A.Vander Lugt, "Signal detection by complex spatial filtering". IEEE Transactions on Information Theory, **10**, 139-145 (1964).
17. I. Kypraios, R. Young, C. Chatwin, "An investigation of the non-linear properties of correlation filter synthesis and neural network
design", Asian Journal of Physics, **Vol. 11, No. 3** (2003).
18. B. V. K. Kumar, "Minimum Variance synthetic discriminant function", Journal of Optical Society of America **A3**, 1579-1584 (1986).
19. A. Mahalanobis, B. V. K. Kumar and D. Casasent, "Minimum average correlation energy filters", Applied Optics, **Vol. 26,No. 17,**
3633-3640 (1987).
20. D. Nguyen and B. Widrow, "The truck backer-upper: An example of self-learning in neural networks, *Proceedings of the International Joint Conference on Neural Networks*, Vol. ***2***, 357-363 (1989).
21. D. Nguyen and B. Widrow, "Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive
weights", *Proceedings of the International Joint Conference on Neural Networks,* Vol. 3, 21-26 (1990).

Performance evaluation for cluttered infrared image based on Fringe-adjusted Joint Transform Correlator

M. S. Alam, E. H. Horache and S. F. Goh
Department of Electrical and Computer Engineering, University of South Alabama
Mobile, Alabama 36688-0002, USA

Abstract: Several metrics for quantifying the performance of fringe-adjusted joint transform correlator (JTC) technique are investigated in this paper. The criteria used for measuring the performance of fringe-adjusted JTC include peak sharpness, signal-to-clutter measure, distortion invariance, signal-to-noise ratio and a new metric called peak-to-background correlation energy is proposed in this paper. These metrics are used to estimate the reliability of signal detection in the input scene with respect to clutter, noise and other associated distortions. Detailed analysis and simulation results for quantifying the performance of fringe-adjusted JTC are presented.

Keywords: Fringe-adjusted joint transform correlator, performance metrics, signal-to-clutter measure, peak-to-sidelobe ratio, peak-to-clutter ratio, signal-to-noise ratio, Fisher ratio, discrimination ratio.

1. Introduction

The joint transform correlation (JTC) technique has been found to be particularly attractive for real time pattern recognition applications [1,2,3,4] such as target detection, fingerprint/face recognition [5,6,7], automatic control of panels [8] and character recognition [9]. The main advantages of the JTC are that it does not require any complex filter fabrication, allows real-time update of the reference image, permits simultaneous Fourier transform of the reference image and the unknown input scene, and relaxes the meticulous alignment required in case of unmatched filter based correlators. The nonlinear JTC [10-11], which uses hardclipping nonlinearity at the Fourier plane yields good correlation performance when compared to the classical JTC in terms of better correlation discrimination, sharper correlation peak intensity, better peak-to-sidelobe ratio (PSR) and narrower correlation width. Recently, Alam et al. proposed the fringe-adjusted JTC [12,13,14], where the joint power spectrum (JPS) is multiplied by a real-valued filter, called fringe adjusted filter (FAF). The FAF is designed to alleviate the problems associated with alternate JTCs, while yielding [16] substantially better correlation output than the classical and binary JTCs for input scenes involving both single and multiple objects.

To measure the performance of a correlator, generally peak sharpness, signal-to-noise ratio and distortion invariance are used [15-16]. These metrics have been used for quantifying the performance of many filter based correlators, such as phase-only filter (POF), binary POF, matched filter, amplitude-modulated POF [17-18] and optimal trade-off filter [19-20] based correlators. In addition, signal-to-clutter measure [21-22] was introduced to measure the difficulty of detecting a target in an unknown input scene. In this paper, we focus on the calculation of aforementioned and related performance metrics to quantify the performance of the fringe-adjusted JTC using challenging forward looking infrared (FLIR) image sequences.

Optical Pattern Recognition XV, edited by David P. Casasent, Tien-Hsin Chao, Proc. of SPIE Vol. 5437 (SPIE, Bellingham, WA, 2004) · 0277-786X/04/$15 · doi: 10.1117/12.548064

2. Performance metrics

Among the various performance metrics reported in the literature, peak-to-sidelobe ratio, signal-to-clutter ratio, peak-to-correlation energy, signal-to-noise ratio, Fisher ratio, and correlation peak height are commonly used to measure the performance of a correlator.

Signal-to-clutter ratio (SCR) is a metric used to determine the level of difficulty posed by clutter that is generally present in the vicinity of the target. It is based on the analysis of the spread of complimentary Eigen values of two classes using Fukunaga-Koontz transform [30,**23**] and is defined by

$$SCR = \sum_{i=N_c+1}^{N} \frac{\lambda_i}{1-\lambda_i} \tag{1}$$

where λ_i represents the eigen value of the target class, and N_c represents the point where the eigen value of target class is higher than the clutter class.

Peak height (PH) is a criterion which measures the height of the correlation peak between the object to be recognized and the reference image, expressed as

$$PH = \left| \left(t(x,y) \otimes r(x,y) \right)_0 \right|^2 \tag{2}$$

where $t(x,y)$ and $r(x,y)$ represent the target and reference image, respectively. Because the correlation is translation invariant, one considers only the central value of the correlation plane.

Peak-to-sidelobe ratio [24] and peak-to-correlation energy are the metrics used to measure the sharpness of the correlation peak, the accurate location of the object to be recognized in the output plane. Thus, it is directly related to the performance of the correlator.

In general, PSR provides the ratio between the correlation peak and the sidelobe. The sidelobe corresponds to a peak at least 2 pixels away from the main peak. The higher the value of PSR the better the correlation performance is. PSR is defined by

$$PSR = \frac{CPI - \mu}{\sigma} \tag{3}$$

where *CPI* represents the correlation peak intensity corresponding to the desired target, μ represents the mean value of the correlation output excluding the peak location (e.g. 3x3 area around the peak), and σ denotes the standard deviation of the correlation output excluding the peak location. Therefore, PSR measures the successful detection of the target.

PCE corresponds to the ratio of the correlation peak intensity with respect to the correlation plane energy, defied as

$$PCE = \frac{CPI}{\sum_y \sum_x \left| c(x,y) \right|^2} \tag{4}$$

where $c(x,y)$ represents the correlation plane. The higher the PCE, the sharper the correlation peak.

When the input scene consists of background noise, stray peaks may appear in the correlation plane. If the amplitude of one of these stray peaks exceeds the desired peak value, it will mislead or complicate the determination of the location of the target. Therefore, it is necessary to know the correlation peak energy over the value of background correlation peaks energy. The corresponding metric, called peak to background correlation energy (PBCE), is defined as

$$PBCE = \frac{CPI}{\sum_y \sum_x \left| b(x,y) \otimes r(x,y) \right|^2} \tag{5}$$

where $\sum_y \sum_x \left| b(x,y) \otimes r(x,y) \right|^2$ represents the correlation energy between the background $b(x,y)$ and reference image, $r(x,y)$. Intuitively, when the correlation energy is concentrated in the peak, the energy of the stray peaks is insignificant.

Signal-to-noise ratio (SNR) provides a measure between signal strength and noise. It is used to estimate the correlator's performance with respect to input scene noise and is defined as

$$SNR = \frac{\sigma^2_{signal}}{\sigma^2_{noise}} \tag{6}$$

where σ^2_{signal} represents the variance of the signal and σ^2_{noise} represents the variance of the noise. A correlator that is less sensitive to noise is preferred because it can produce relatively similar output in the presence of noise.

Peak-to-clutter ratio (PCR) provides a measure involving the correlation peaks of the desired target and the background, defined as

$$PCR = \frac{Peak_{target}}{Peak_{clutter}} \tag{7}$$

where $Peak_{target}$ is the peak value corresponding to the target, and $Peak_{clutter}$ is the peak value corresponding to clutter. In general, if PCR is greater than 1, the correlator will succeed in identifying the target. Otherwise, it may fail.

Fisher Ratio (FR) is used as a measure to evaluate the performance of a correlator for distortion invariant target detection, defined as

$$FR = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2} \tag{8}$$

where μ_1 and σ_1^2 represent the mean and variance of the correlation plane for the previous frame, while μ_2 and σ_2^2 represent the mean and variance of the correlation plane for the current frame, respectively.

The discrimination performance metric is defined as

$$D = 100\frac{P_1 - P_2}{P_1} \tag{9}$$

where P_1 is the desired peak value and P_2 is the maximum of the stray peak value resulting from various factors such as background noise, clutter and other artifacts.

Discrimination Ratio (DR) corresponds to the discrimination between two classes of objects, such as class A and class B, where class A corresponds to the desired target(s), class B corresponds to the non-target object(s). This performance metric is defined as

$$DR = \frac{\min_{x,y\in target set}\left\{\left|c(x,y)\right|^2\right\}}{\max_{x,y\in anti-target set}\left\{\left|c(x,y)\right|^2\right\}} \tag{10}$$

where $\max_{x,y\in anti-target set}\left\{\left|c(x,y)\right|^2\right\}$ indicates the maximum peak value of non-target object, and $\min_{x,y\in target set}\left\{\left|c(x,y)\right|^2\right\}$ represents the minimum peak value of the desired target. If DR is greater than 1, class A can be detected successfully from the distorted scene. Otherwise, it may fail.

3. Analysis

Assume that the reference and the input scene used in a fringe-adjusted JTC [12 -13] is expressed as $r(x,y+y_0)$ and $s(x,y-y_0)$ respectively. Because the input scene $s(x,y-y_0)$ consists of the target, $t(x,y-y_0)$, and background, $b(x,y-y_0)$, the input joint image can be expressed as

$$f(x,y) = r(x,y+y_0) + t(x,y-y_0) + b(x,y-y_0) \tag{11}$$

where $b(x,y) = b_0(x,y)\left[1 - w(x,y)\right]$ and w(x,y) is the target window. Applying Fourier transform to Eq. (11) yields,

$$F(u,v) = \left|R(u,v)\right|e^{j\phi_R}e^{jvy_0} + \left|T(u,v)\right|e^{j\phi_T}e^{-jvy_0} + \left|B(u,v)\right|e^{j\phi_B}e^{-jvy_0} \tag{12}$$

The joint power spectrum corresponding to Eq. (12) may be obtained as

$$\begin{aligned} I(u,v) = {} & \left|R(u,v)\right|^2 + \left|T(u,v)\right|^2 + \left|B(u,v)\right|^2 + 2\left|R(u,v)\right|\left|T(u,v)\right| \\ & \times\cos(\phi_R(u,v) - \phi_T(u,v) + 2vy_0) + 2\left|R(u,v)\right|\left|B(u,v)\right| \\ & \times\cos(\phi_R(u,v) - \phi_B(u,v) + 2vy_0) + 2\left|T(u,v)\right|\left|B(u,v)\right| \\ & \times\cos(\phi_T(u,v) - \phi_B(u,v)) \end{aligned} \tag{13}$$

The fringed adjusted filter (FAF) used in a fringe-adjusted JTC is defined as $FAF = \frac{C(u,v)}{D(u,v) + \left|R(u,v)\right|^2}$, where

$C(u,v)$ and $D(u,v)$ are either constants or functions of u and v. Thus, the fringe-adjusted JPS may be expressed as

$$
\begin{aligned}
G(u,v) = \frac{C(u,v)}{D(u,v)+|R(u,v)|^2} \times \{ & |R(u,v)|^2 + |T(u,v)|^2 + |B(u,v)|^2 \\
& + 2|R(u,v)||T(u,v)|\cos(\phi_R(u,v) - \phi_T(u,v) + 2vy_0) \\
& + 2|R(u,v)||B(u,v)|\cos(\phi_R(u,v) - \phi_B(u,v) + 2vy_0) \\
& + 2|T(u,v)||B(u,v)|\cos(\phi_T(u,v) - \phi_B(u,v))\}
\end{aligned}
\tag{14}
$$

The inverse Fourier transform for Eq. (14) yields the correlation output. The first three terms corresponds to the zero-order term, the fourth term corresponds to the crosscorrelation between the target and the reference, the fifth term corresponds to the crosscorrelation between the background and the reference image and the sixth term corresponds to the crosscorrelation between the target and background.

4. Simulation Results

In this section, we investigate the performance of the fringe-adjusted JTC with respect to the previously mentioned performance metrics. For testing purposes, we considered the detection of a target in a FLIR image sequence consisting of 290 frames (Sequence #1). Figure 1 shows five sample frames of this sequence.

From Figure 2, it is evident that the SCR from frame #60 and onwards is almost identical indicating that there is distinct difference between the target and clutter. It dips around frame #40 where the target blends slightly with the background. However, it did not pose any difficulty for fringe-adjusted JTC to detect the target. For comparison, we plot the SCR for another sequence (Sequence #2), which contains an unclear target as shown in Fig. 3. Although the SCR value appears to be lower than the first sequence, the fringe-adjusted JTC successfully detects the target.

Figure 4 shows the target peak height for the input scene with and without considering the effects of background. Average light efficiency for all frames of Sequence #1 is 0.23 without_considering the effects of background, and 0.02 considering the effects of background. This is due to the fact that the energy is divided between the fifth and sixth terms of Eq. (14) in the correlation output.

Figure 5 shows the PSR for each frame of Sequence #1. From Fig. 5, it is evident that the PSR gradually increases because the target size increases and becomes clearer with the increase in frame number. Figure 6 shows the relationship between the PSR, target size, and contrast of the target for sequence #1.

Figure 7 shows the PCE for each frame of Sequence #1 indicating that the fringe-adjusted JTC yields sharp correlation peak. On an average, 4.5% of the correlation energy is concentrated in the correlation peak. Figure 8 shows the PBCE for all frames of Sequence #1. From Fig 8, it is obvious that the PBCE varies with changes in target background in the input scene.

Figure 9 depicts the PCR for all frames of Sequence #1. From Figure 9, we observe that the PCR is greater than 1 for all frames of the sequence implying that the correlation peak corresponding to the target is higher than the stray peaks. As a result, it posed no difficulty in detecting the target by locating the maximum peak in the correlation plane.

Figure 10 shows a plot of PSR versus SNR for all frames of sequence #1. With SNR, we can measure the

strength of the signal with respect to noise. From Figure 10, it is evident that when PSR increases with the increase in SNR, and the target can be easily detected with the reduction in input scene noise.

The Fisher ratio (FR) is used to show the similarity and dissimilarity between the current and preceding frame of a sequence. The FR involving the current and preceding frames for all frames of Sequence #1 are shown in Figure 11. For example, the FR is higher between frames 77 and 78 of Sequence #1 indicating that these two frames are highly dissimilar.

Figure 12 shows the discrimination capability of the fringe-adjusted JTC between the target and background. Figure 12 depicts that when the discrimination ratio (DR) is higher, it is easy to detect the target from the background, otherwise it is difficult to detect.

Finally, we rotated the target by 20 degrees at 1° interval to generate an SDF image, which was used as the reference image. Then, various rotated versions of the target were used as the input scene in the fringe-adjusted JTC to obtain the correlation output. Thereafter, the input scene containing only the background was introduced into the fringe-adjusted JTC to obtain the correlation output with respect to the background. Figure 13 shows the plot of DR for all frames of Sequence #1. From Figure 13, it is obvious that the DR is less than 1 for the first 200 frames of Sequence #1. This is due to the small target size and that the SDF created cannot recognize all of the rotated versions of the target. However, when the target size is increased, the DR increases as well. When DR is greater than 1, it is evident from Fig. 13 that all rotated target can be recognized.

5. Conclusion

We investigated the performance of fringe-adjusted JTC using various metrics. Peak height, peak sharpness, signal-to-noise ratio, peak-to-clutter ratio, discriminability and distortion invariance are used to measure the performance for fringe-adjusted JTC. From the analysis, we conclude that the size and contrast of the target are very important for detection purposes. If the size of the target is very small, such as 5 pixels, and the contrast is similar to the background, it is difficult to detect the target. On the other hand, a small target with high contrast or a large target with low contrast can be easily detected. When the input scene is corrupted by noise, it degrades the performance of the correlator. Thus, image preprocessing becomes necessary to alleviate the effects of noise and other detrimental factors before applying the fringe-adjusted JTC algorithm.

Acknowledgement

This research was supported by a grant from the Army Research Office (DAAD 19-02-1-0150). The authors wish to thank Richard Sims and William Sanders for many rewarding discussions.

References

1. C. S. Weaver and J. W. Goodman, "A technique for optically convolving two functions;" appl. Opt., vol. 5, pp. 1248-1249, 1966.
2. F. T. S. Yu and X. J. Lu, "A programmable joint transform correlator," Opt. Comm. 52, 10-16, (1984).
3. J. Wang and B. Javidi, "Multiobject detection using the binary joint transform correlator with different types of thresholding methods," Opt. Eng. 33, 1793-1804 (1994).
4. M. Deutsch, J. Garcia, and D. Mendlovic, "Multichannel single-output color pattern recognition by use of a joint-transform correlator," Appl. Opt. 35, 6976-6982 (1996).

5. P. K. H. Fielding, J. L. Horner, and C. K. Makekau, "Optical fingerprint identification by binary joint transform correlation," Opt. Eng. 30, 1958-1961 (1991).
6. J. Rodolfo, H. Rajbenbach, and J.-P. Huignard, "Performance of a photorefractive joint transform correlator for fingerprint identification," Opt. Eng. 34, 1166-1171 (1995).
7. T. Grycewicz and B. Javidi, "Experimental comparison of binary joint transform correlators used for fingerprint identification," Opt. Eng. 35, 2519-2525 (1996).
8. L. Guibert, G. Keryer, A. Servel; M. Attia, M S. MacKenzie, P. Pellat-Finet and J. L. de-Bougrenet-de-la-Tocnaye, "On-board optical joint transform correlator for real-time road sign recognition Opt. Eng. vol.34, no.1; Jan. 1995; p.1/135-43.
9. T. Iwaki, Y. Mistuoka, "Optical pattern recognition of letters by a joint-transform correlator using a ferroelectric liquid-crystal spatial light modulator", Optics Letters, Volume 15, Issue 21, 1218, November 1990.
10. B. Javidi, "Nonlinear Joint Power Spectrum Based Optical Correlation," Journal of Applied Optics, Vol. 28, No. 12, pp. 2358-2367, June 15, 1989.
11. B. Javidi and J. L. Horner, "Single Spatial Light Modulator Binary Nonlinear Optical Correlator," Proceedings of SPIE on Optical Pattern Recognition, Vol. 1014, Symposium on Innovative Science and Technology, Los Angeles, California, January 1989.
12. M. S. Alam and M. A. Karim, "Fringe-adjusted joint transform correlator," Appl. Opt. 32, p. 4344-4350 (1993).
13. M. S. Alam and M. A. Karim, "Joint-transform correlator under varying illumination," Appl. Opt. 32, p. 4344-4350 (1993).
14. M. S. Alam, "Performance of Joint Transform Correlation Under Varying Illumination and Noise," Journal of Optics and Laser Technology, Vol. 27, p. 361-368, 1995
15. B. V. K. Vijaya Kumar and L. Hassebrook, "Performance measures for correlation filters", Applied-Optics. vol.29, no.20; 10 July 1990; p.2997-3006.
16. J. L. Horner, ''Metrics for assessing pattern recognition performance,''Appl. Opt. 31, 165–166 119922.
17. S. H. Zheng, M. A. Karin and M. L. Gao, "Amplitude-modulated phase-only filter performance in presence of noise," Optics Communition, 89, 296-305, (1992)
18. D. C. Wunsch, R. J. Marks, T. P. Caudell, and C. D. Capps,"Limitations of a class of binary phase-only filters,"APPLIED OPTICS / Vol. 31, No. 26 / 10 September 1992.
19. B. V. K. Vijaya Kumar, D. W. Carlson, and A. Mahalanobis,"Optimal trade-off synthetic discriminant function filters for arbitrary devices," Opt. Lett. 19, 1556–1558 (1994).
20. P.. Réfrégier, "Optimal trade-off filters for noise robustness,sharpness of the correlation peak, and Horner efficiency," Opt.Lett. 16, 829–831 (1991).
21. M. A. Philips, S. R. F. Sims, " A signal to clutter measure for ATR performance comparison," In proceedings of SPIE Automatic Target Recognition VII, pp. 74-81, Vol. 3069, April 1997.
22. A. Mahalanobis, S. R. F. Sims, A.V. Nevel, "Signal-to-clutter measure for measuring automatic target recognition performance using complimentary eigenvalue distribution analysis," Opt. Eng. 42,1144-1151, (2003)
23. K. Fukunaga and W.L.G. Koontz, "Representation of random processes using the finite Karhunen-Loeve expansion," IEEE Trans. Info. 16(1), 85-10 (1970).
24. R. Singh and B.V.K. Vijaya Kumar,"Performance of the extended maximum average correlation height (EMACH) filter and the polynomial distance classifier correlation filter (PDCCF) for multi-class SAR detection and classification," Proceedings. Of SPIE, Vol. 4727, No. 23, (2002).

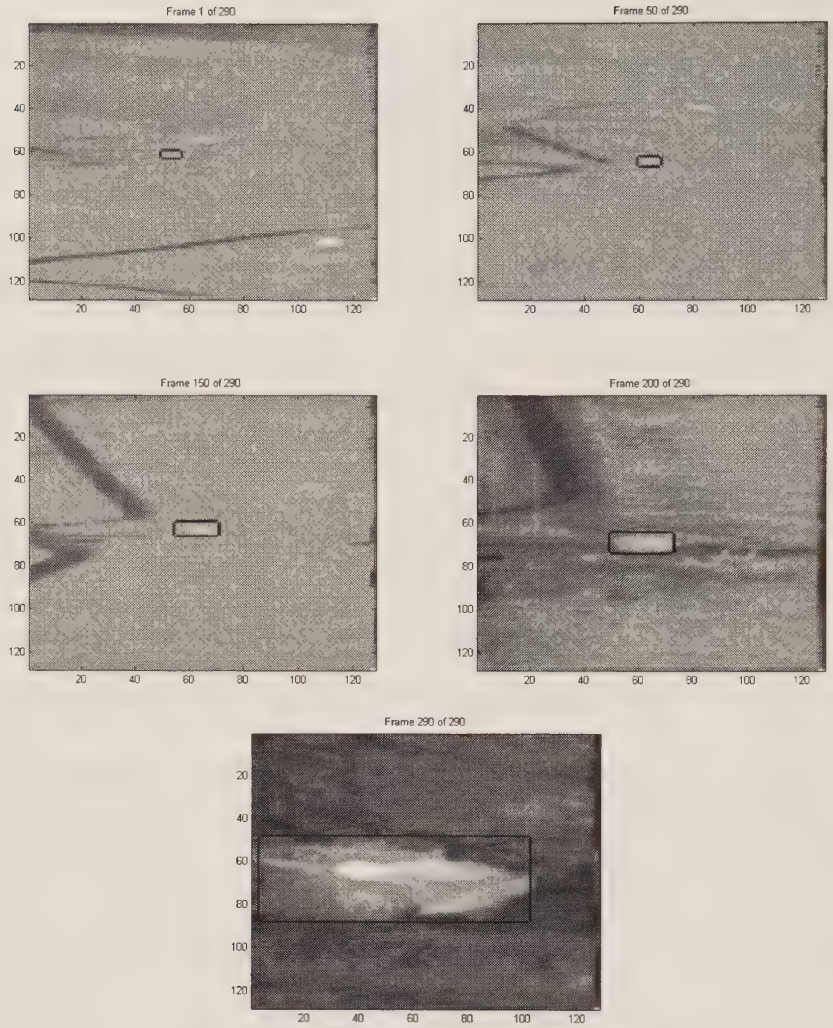


Figure 1: 5 out of 290 frames of FLIR sequence #1 (all frame are 128x128 pixels).

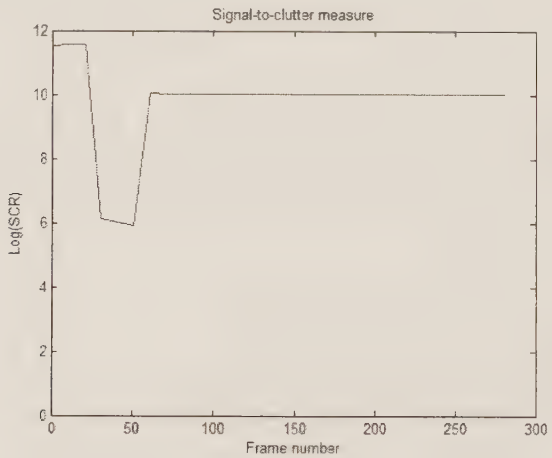


Figure 2: Signal-to-clutter ratio for all frames of sequence #1.

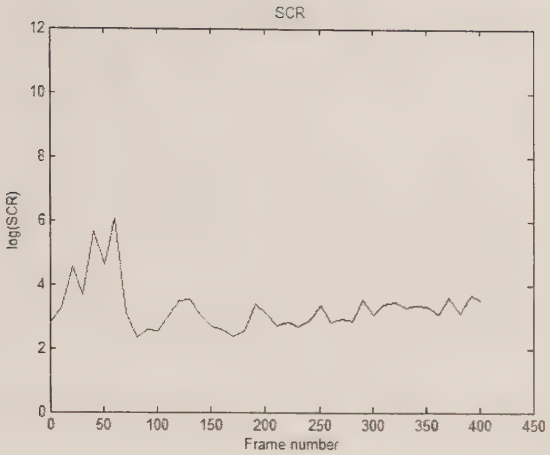


Figure 3: SCM for sequence #2 for comparison.

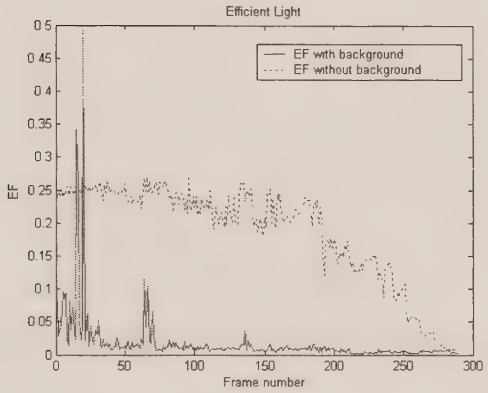


Figure 4: Peak height for all frames of sequence #1 with and without considering the effects of background.

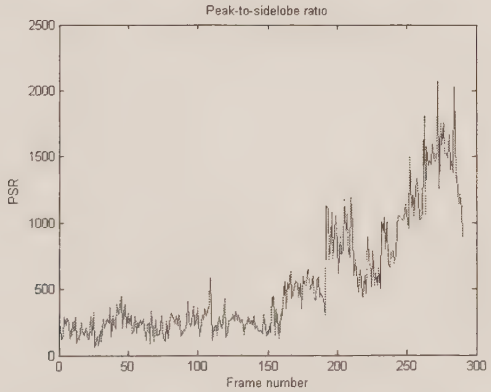


Figure 5: Peak-sidelobe-ratio for each frame of sequence #1.

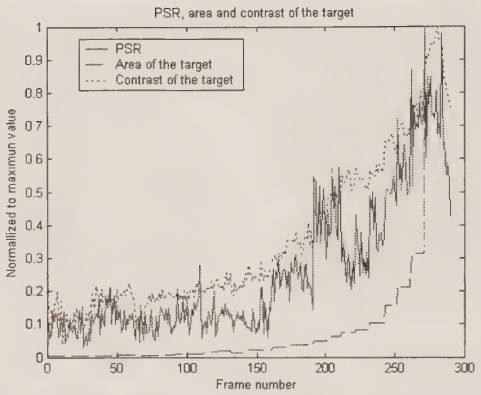


Figure 6: Peak-to-sidelobe ratio, contrast and area of the target.

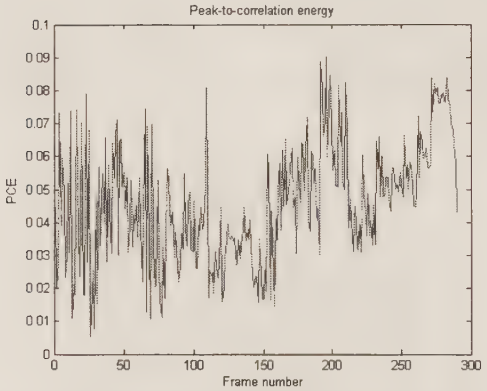


Figure 7: Peak-to-Correlation Energy (PCE) for all frames of sequence #1.

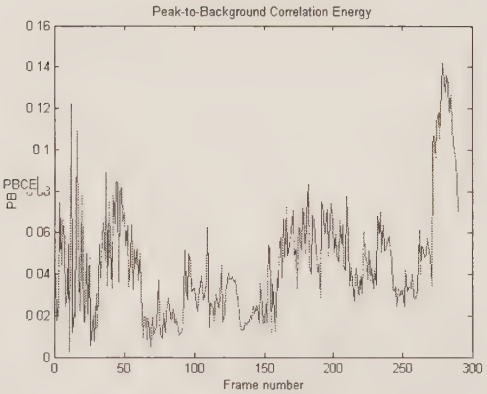


Figure 8: Peak-to-Background Correlation Energy (PBCE) for all frames of sequence #1.

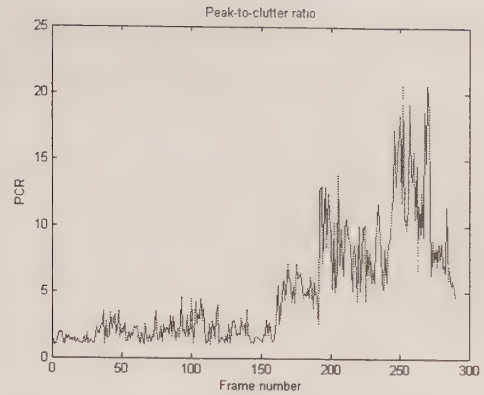


Figure 9: Peak-to-clutter ratio for all frames of sequence #1.

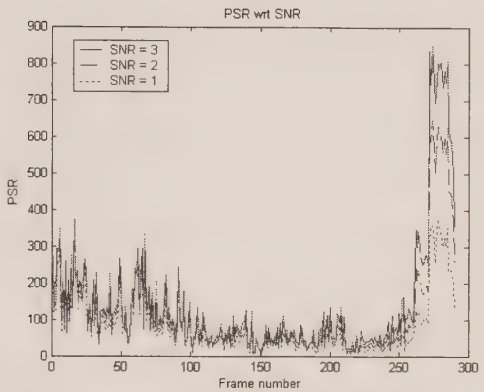


Figure 10: PSR versus SNR for all frames of sequence #1.

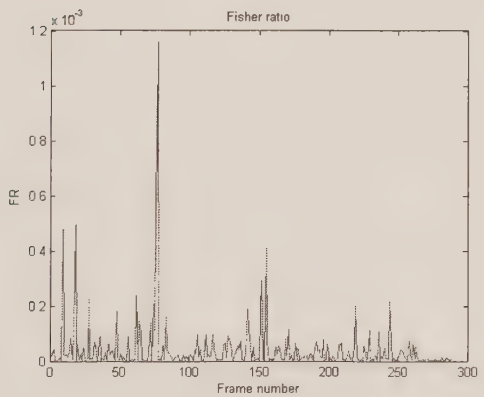


Figure 11: Fisher Ratio between the current and preceding frame for sequence #1.

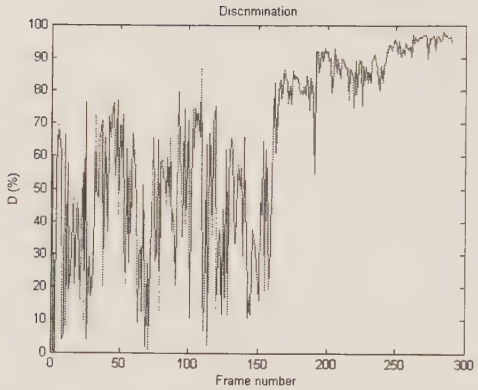


Figure 12: Discrimination capability of the fringe-adjusted JTC between the target and background for sequence #1.

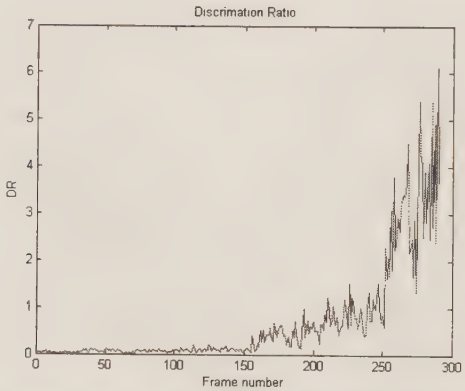


Figure 13: Discrimination ratio, the target was rotated 20 degrees at 1° interval to generate SDF image, which was used as the reference image.

Modified phase-encoded fringe-adjusted joint transform correlation for multiple target detection

M. R. Haider[a], M. N. Islam[a], M. S. Alam[b] and J. F. Khan[*b]
[a]Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology, Dhaka-1000, Bangladesh
[b]Department of Electrical and Computer Engineering
University of South Alabama, Mobile, AL 36688-0002, USA

ABSTRACT

A modified phase-encoded fringe-adjusted joint transform correlation technique is proposed for multiple target detection, where two joint power spectrums (JPS) are formulated utilizing a random phase mask and phase shifted random phase mask to the reference image separately. The final JPS is the difference between the phase encoded and shifted phase encoded JPS which is multiplied by the phase mask before applying the inverse Fourier transform to yield the correlation output. This technique ensures better utilization of the input/output plane space bandwidth product by yielding one delta function like correlation peak for each desired target object and no peak for non-target objects. The proposed technique can effectively detect any number of targets from noise free or noisy input scenes without changing the system parameters and without any degradation of performance. Computer simulation results verify the performance of the proposed technique.

Keywords: Joint transform correlator, fringe-adjusted filter, phase mask, phase encoding, space bandwidth product, joint power spectrum.

1 INTRODUCTION

Correlation based optical pattern recognition is usually performed by using the ability of a lens to perform two-dimensional Fourier transform operation. Optical pattern recognition techniques are generally implemented using the VanderLugt filter based correlator[1] or the joint transform correlator (JTC)[2]. A VanderLugt correlator requires a priori fabrication of a complex matched filter which is used in the Fourier plane. Moreover, this filter must be accurately aligned along the optical axis and requires intimate positioning between the filter and the Fourier transform of the input, which complicates the real time operation of the VanderLugt correlator. On the other hand, a JTC can be operated at video frame rates[3], and does not require a priori fabrication of a complex filter and meticulous alignment along the optical axis. Although a classical JTC provides attractive advantages over matched filter-based correlators[4], it suffers from low diffraction efficiency, broad correlation width and discrimination sensitivity[5-9]. A binary JTC has been found to be superior to a classical JTC[5,9]. However, a binary JTC involves Fourier plane joint power spectrum (JPS) binarization, which may limit the system processing speed[10]. For multi-object input scenes, in particular, the JPS binarization process introduces harmonic correlation peaks and thereby complicates the target detection process[9,10]. Recently, a phase-shifting technique has been effectively utilized in a JTC to obtain non-zero order correlation output[11,12]. The fringe-adjusted JTC technique has been found to alleviate the problems associated with alternate JTC techniques[13,14]. In the fringe-adjusted JTC, the JPS is multiplied by the real valued fringe-adjusted filter (FAF) at the Fourier plane before applying the inverse Fourier transform to yield the correlation output.

For multiple target detection, several multi-step JTC algorithms involving Fourier plane image subtraction[15], correlation plane image subtraction[16] and phase-shifted power spectrum subtraction[17] had been proposed. These algorithms involve multiple processing steps and yield better correlation output by suppressing the zero-order term and false alarms due to autocorrelation among similar input scene target and non-target objects.

*jkhan@jaguar1.usouthal.edu; phone 1 251 460-7484; fax 1 251 460-6028

Optical Pattern Recognition XV, edited by David P. Casasent, Tien-Hsin Chao, Proc. of SPIE Vol. 5437 (SPIE, Bellingham, WA, 2004) · 0277-786X/04/$15 · doi: 10.1117/12.548065

The above mentioned JTCs generally utilize 50% of the input plane spatial light modulator (SLM) to display the input scene and 50% of the output plane to display the correlation output thus making poor utilization of the space bandwidth product (SBP). Recently, a phase encoded fringe-adjusted JTC (PJTC) technique has been proposed, which can detect multiple targets in one processing step and ensures better utilization of the SBP by generating one correlation peak per target[18]. However, for multi-target noisy input scenes, the PJTC technique may require the adjustment of a parameter used in the FAF formulation for successful detection of a target. To overcome the aforementioned problem and, at the same time, to enhance the correlation peak intensity, a phase-shifted PJTC (SPJTC) technique is proposed in this paper, which involves the phase-shifted power spectrum subtraction technique and the PJTC technique. Simulation results show that the proposed SPJTC technique yields single correlation peak per target with high discrimination capability and better noise robustness.

2 ANALYSIS

In a classical JTC architecture the reference image and the input image are displayed side by side using a spatial light modulator (SLM) illuminated by a coherent light source. If $r(x, y+y_0)$ represents the reference image and $t(x, y-y_0)$ represents the input scene containing n objects $t_1(x-x_1, y-y_1)$, $t_2(x-x_2, y-y_2)$, ... $t_n(x-x_n, y-y_n)$, the input joint image may be expressed as

$$f(x, y) = r(x, y+y_0) + \sum_{i=1}^{n} t_i(x-x_i, y-y_i) \tag{1}$$

Applying Fourier transform to the input joint image by the lens L1 yields

$$F(u,v) = |R(u,v)| \exp[j\Phi_r(u,v) + jvy_0] + \sum_{i=1}^{n} |T_i(u,v)| \exp[j\Phi_{ti}(u,v) + jux_i - jvy_i] \tag{2}$$

where $R(u,v)$ and $T_i(u,v)$ are the amplitudes, $\Phi_r(u,v)$ and $\Phi_{ti}(u,v)$ are the phases of the Fourier transforms of $r(x, y)$ and $t_i(x, y)$, respectively; u and v are mutually independent frequency domain variables scaled by a factor of $\frac{2\pi}{\lambda f}$, λ is the wavelength of collimating light, f is the focal length of lenses. The complex light distribution produced in the back focal plane of one lens, called the JPS, is then detected by a square-law detector such as a CCD array. The JPS is given by

$$\begin{aligned} |F(u,v)|^2 = {} & |R(u,v)|^2 + \sum_{i=1}^{n} |T_i(u,v)|^2 + 2\sum_{i=1}^{n} |T_i(u,v)| \, |R(u,v)| \cos[\Phi_{ti}(u,v) - \Phi_r(u,v) \\ & - ux_i - vy_i - 2vy_0] + 2\sum_{i=1}^{n} \sum_{k=1, k\neq i}^{n} |T_i(u,v)| \, |T_k(u,v)| \cos[\Phi_{ti}(u,v) - \Phi_{tk}(u,v) - ux_i + ux_k - vy_i + vy_k] \end{aligned} \tag{3}$$

From Eq. (3), it is evident that the correlation output contains autocorrelation of the reference image and the input scene objects, cross-correlation between the reference image and the input scene objects, and cross-correlations between different input scene objects. The first, second and fourth terms of Eq. (3) produce undesired correlation peaks, and only the third term generates the desired correlation peaks. The undesired correlation peaks can be eliminated by using Fourier plane image subtraction[15], correlation plane image subtraction[16], or phase-encoded fringe-adjusted JTC[18]. In a classical JTC, the JPS of Eq. (3) is introduced into a second SLM, which is illuminated by the same laser using a beam splitter and mirror combination. In a fringe-adjusted JTC technique, the JPS is multiplied by the FAF transfer function before applying the inverse Fourier transform operation. The FAF transfer function[13] is defined as

$$H(u,v) = \frac{C(u,v)}{D(u,v) + |R(u,v)|^2} \tag{4}$$

where $C(u, v)$ and $D(u, v)$ are either constants or functions of u and v. The parameter $C(u,v)$ is used to avoid having an optical gain greater than unity, while $D(u,v)$ is used to overcome the pole problem otherwise associated with a normal

inverse filter. The fringe-adjusted JPS is obtained by multiplying the JPS of Eq. (3) by the filter transfer function *H(u,v)* as follows:

$$G(u,v) = H(u,v) \times |F(u,v)|^2 \tag{5}$$

An inverse Fourier transform of the fringe-adjusted JPS performed by a lens yields the correlation output.

For multiple target detection, either Fourier plane image subtraction[15], or correlation plane image subtraction[16], or phase-encoded fringe-adjusted JTC[18] may be used depending on the system constraint. The Fourier plane and correlation plane image subtraction techniques involve multiple processing steps and generate a pair of correlation peaks for each target. The PJTC technique overcomes these problems by yielding one peak per target in one processing step. In the PJTC technique, a random phase mask, defined by

$$\Phi(u,v) = \exp[j\Psi(u,v)] \tag{6}$$

is multiplied by the input scene power spectrum, yielding

$$|T_{ph}(u,v)| = \sum_{i=1}^{n} |T_i(u,v)| \exp[j\Phi_{ti} - j(ux_i - vy_i)] \times \Phi(u,v) \tag{7}$$

An inverse Fourier transform of Eq. (7) yields the phase encoded input scene, given by

$$t_{ph}(x,y) = \sum_{i=1}^{n} t_i(x - x_i, y - y_i) \otimes \phi(x,y) \tag{8}$$

where $\otimes$ represents convolution operation. The phase-encoded input scene of Eq. (8) is combined with the reference image to form the input joint image, which is then Fourier transformed to yield the phase-encoded JPS, given by

$$\begin{aligned}
|F(u,v)|^2 &= |R(u,v)|^2 + \sum_{i=1}^{n} |T_i(u,v)|^2 \\
&+ \sum_{i=1}^{n} |R(u,v)||T_i(u,v)| \exp[j(\Phi_r(u,v) - \Phi_{ti}(u,v) + ux_i + v(y_0 + y_i))] \times \Phi^*(u,v) \\
&+ \sum_{i=1}^{n} |R(u,v)||T_i(u,v)| \exp[-j(\Phi_r(u,v) - \Phi_{ti}(u,v) + ux_i + v(y_0 + y_i))] \times \Phi(u,v) \\
&+ 2\sum_{i=1}^{n} \sum_{k=1, k\neq i}^{n} |T_i(u,v)||T_k(u,v)| \cos[\Phi_{ti}(u,v) - \Phi_{tk}(u,v) - u(x_i - x_k) - v(y_i - y_k)]
\end{aligned} \tag{9}$$

where the superscript * represents the complex conjugate. The phase-encoded JPS is again multiplied by the same phase mask $\Phi(u,v)$ to yield the modified JPS, given by

$$\begin{aligned}
O(u,v) &= |R(u,v)|^2 \Phi(u,v) + \sum_{i=1}^{n} |T_i(u,v)|^2 \Phi(u,v) \\
&+ \sum_{i=1}^{n} |R(u,v)||T_i(u,v)| \exp[j(\Phi_r(u,v) - \Phi_{ti}(u,v) + ux_i + v(y_0 + y_i))] \\
&+ \sum_{i=1}^{n} |R(u,v)||T_i(u,v)| \exp[-(\Phi_r(u,v) - \Phi_{ti}(u,v) + ux_i + v(y_0 + y_i))] \times \Phi(u,v)\Phi(u,v) \\
&+ 2\sum_{i=1}^{n} \sum_{k=1, k\neq i}^{n} |T_i(u,v)||T_k(u,v)| \cos[\Phi_{ti}(u,v) - \Phi_{tk}(u,v) - u(x_i - x_k) - v(y_i - y_k)]\Phi(u,v)
\end{aligned} \tag{10}$$

It is obvious from Eq. (10) that the third term, which yields the desired correlation peak, is not affected by the phase mask, and the unwanted or extraneous correlation signals are intentionally disturbed by the phase mask. If the phase mask is chosen carefully, the extraneous correlation signals can be scattered in various directions so that the zero-order terms and false alarms can automatically be eliminated. Also, only one correlation peak is generated for each target[18].

To achieve high discrimination capability, the phase-encoded JPS is multiplied by FAF given by

$$O_{faf}(u,v) = \frac{C(u,v) \times O(u,v)}{D(u,v) + |R(u,v)|^2} \tag{11}$$

However, it is observed that the PJTC technique is sensitive to noise and the number of target and non-target objects present in the input scene. The aforementioned problem can be tackled by adjusting the FAF parameter, $D(u,v)$, which is not always a convenient option for practical applications. With the increase of targets in a noisy input scene, the amplitude of noise peaks in the correlation plane surface start to rise which complicates the detection process. Assume that the input scene contain two identical targets, $t_1(x-x_1,y-y_1)$ and $t_2(x-x_2,y-y_2)$, which are similar to the reference image $r(x, y)$. Then the modified JPS of Eq. (11) can be expressed as

$$O_{faf}(u,v) = \frac{C(u,v)}{D(u,v) + |R(u,v)|^2} \times \begin{bmatrix} |R(u,v)|^2 \, \Phi(u,v) + 2|R(u,v)|^2 \, \Phi(u,v) \\ + \sum_{i=1}^{2} |R(u,v)|^2 \exp[j(ux_i + v(y_0 + y_i))] \\ + \sum_{i=1}^{2} |R(u,v)|^2 \exp[-j(ux_i + v(y_0 + y_i))] \times \Phi(u,v) \times \Phi(u,v) \\ + 2\sum_{i=1}^{2} \sum_{k=1,k\neq i}^{2} |R(u,v)|^2 \cos[-ux_i - ux_k - vy_i - vy_k] \times \Phi(u,v) \end{bmatrix} \tag{12}$$

From Eq. (12), it is evident that all terms involve constant magnitude spectra, and only the third term produces a sharp correlation peak while the remaining correlation peaks are steered out of the correlation plane due to the application of the phase mask.

Now considering a case involving one target $t_1(x-x_1,y-y_1)$ and a non-target object $t_2(x-x_2,y-y_2)$ Eq. (12) can be written as

$$\begin{aligned} O_{faf}(u,v) = {} & \frac{C(u,v)}{D(u,v) + |R(u,v)|^2} \times [2|R(u,v)|^2 \times \Phi(u,v) \\ & + |T_2(u,v)|^2 \times \Phi(u,v) + |R(u,v)|^2 \exp[j(ux_1 + v(y_0 + y_1))] \\ & + |R(u,v)|^2 \exp[-j(ux_1 + v(y_0 + y_1))] \times \Phi(u,v) \times \Phi(u,v) \\ & + |T_2(u,v)| \times |R(u,v)| \exp[j(\Phi_r(u,v) - \Phi_{t2}(u,v) + ux_2 + v(y_0 + y_2))] \\ & + |T_2(u,v)| \times |R(u,v)| \exp[-j(\Phi_r(u,v) - \Phi_{t2}(u,v) + ux_2 + v(y_0 + y_2))] \times \Phi(u,v) \times \Phi(u,v) \\ & + 2\sum_{i=1}^{2} \sum_{k=1,k\neq i}^{2} T_i(u,v) \| T_k(u,v) | \cos[\Phi_{ti}(u,v) - \Phi_{tk}(u,v) - ux_i + ux_k - vy_i + vy_k] \times \Phi(u,v)] \end{aligned} \tag{13}$$

In Eq. (13), only the first, third and fourth terms contain constant magnitude spectra, while the remaining terms contain magnitude spectra amplitude-modulated by the corresponding non-target object. Therefore, $D(u,v)$ must be selected to ensure that the desired correlation peak is significantly higher than the false alarms.

To overcome the limitations of the PJTC technique, a new technique, called shifted phase-encoded fringe-adjusted JTC (SPJTC), is proposed in this paper. Here, the reference image is Fourier transformed and fed to two channels, as shown in Fig. 1, where one channel introduces 180^0 phase shift. Then a random phase mask is applied to both channels and the resultant signals are given by

$$S_1(u,v) = |R(u,v)| \exp[j\Phi_r(u,v) + jvy_0] \times \Phi(u,v) \tag{14}$$

$$S_2(u,v) = |R(u,v)| \exp[j\pi] \exp[j\Phi_r(u,v) + jvy_0] \times \Phi(u,v) \tag{15}$$

Applying inverse Fourier transform to Eqs. (14) and (15), we get

$$s_1(x,y) = r(x, y + y_0) \otimes \phi(x,y) \tag{16}$$

$$s_2(x,y) = -r(x, y + y_0) \otimes \phi(x,y) \tag{17}$$

The input scene is then added to both channels to form two input joint images, given by

$$f_1(x,y) = r(x, y+y_0) \otimes \phi(x,y) + \sum_{i=1}^{n} t_i(x, y-y_i) \tag{18}$$

$$f_2(x,y) = -r(x, y+y_0) \otimes \phi(x,y) + \sum_{i=1}^{n} t_i(x, y-y_i) \tag{19}$$

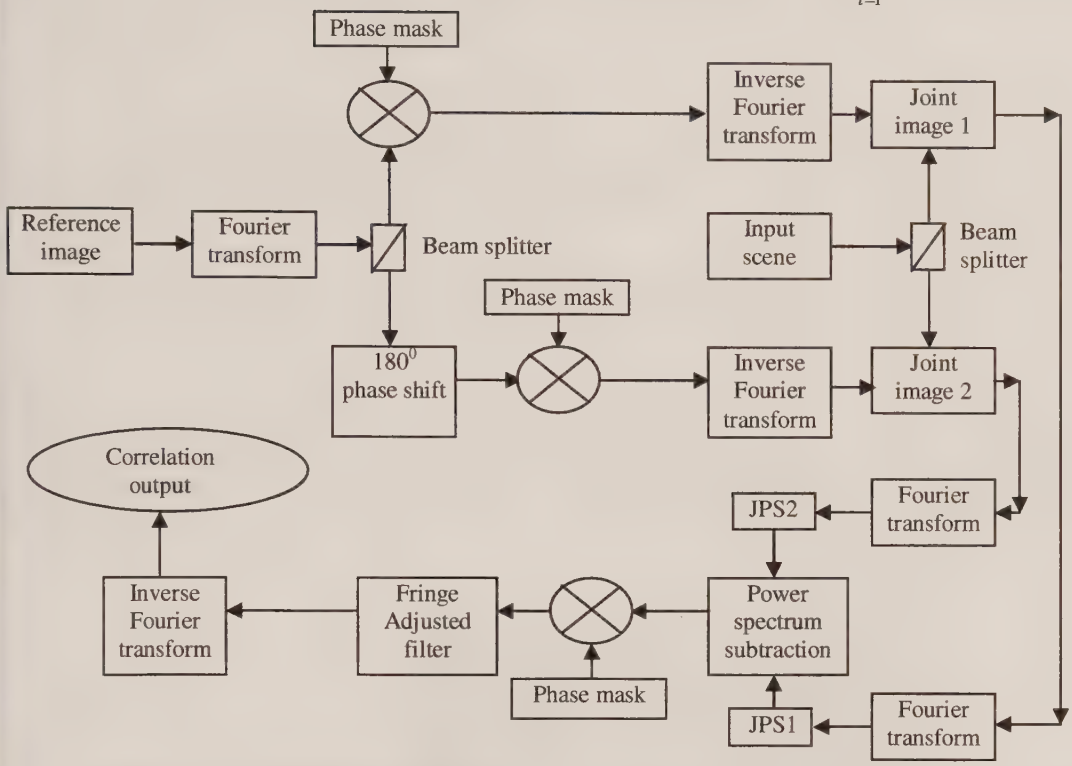


Fig. 1. Architecture of the proposed SPJTC scheme

The 180^0 phase-shifted version of the reference image, $s_2(x, y)$, can be displayed in a SLM by adding a dc value as suggested by Yu et. al.[11]. Applying Fourier transform to Eqs. (18) and (19), we get

$$\begin{aligned} |F_1(u,v)|^2 = {} & |R(u,v)|^2 + \sum_{i=1}^{n} |T(u,v)|^2 + \sum_{i=1}^{n} R(u,v) T_i^*(u,v) \exp[j(vy_0 - vy_i)] \times \Phi(u,v) \\ & + \sum_{i=1}^{n} R^*(u,v) T_i(u,v) \exp[-j(vy_0 - vy_i)] \times \Phi^*(u,v) + \sum_{i=1}^{n} \sum_{k=1,k\neq i}^{n} T_i(u,v) T_k^*(u,v) \exp[-j(vy_i - vy_k)] \\ & + \sum_{i=1}^{n} \sum_{k=1,k\neq i}^{n} T_i^*(u,v) T_k(u,v) \exp[j(vy_i - vy_k)] \end{aligned} \tag{20}$$

and

$$\begin{aligned} |F_2(u,v)|^2 = {} & |R(u,v)|^2 + \sum_{i=1}^{n} |T(u,v)|^2 - \sum_{i=1}^{n} R(u,v) T_i^*(u,v) \exp[j(vy_0 - vy_i)] \times \Phi(u,v) \\ & - \sum_{i=1}^{n} R^*(u,v) T_i(u,v) \exp[-j(vy_0 - vy_i)] \times \Phi^*(u,v) + \sum_{i=1}^{n} \sum_{k=1,k\neq i}^{n} T_i(u,v) T_k^*(u,v) \exp[-j(vy_i - vy_k)] \\ & + \sum_{i=1}^{n} \sum_{k=1,k\neq i}^{n} T_i^*(u,v) T_k(u,v) \exp[j(vy_i - vy_k)] \end{aligned} \tag{21}$$

The JPS corresponding to Eqs. (20) and (21) can be calculated in parallel and does not have detrimental effect on the processing speed. The phase-encoded JPS of Eq. (20) is then subtracted from that of Eq. (21) and the corresponding output is multiplied by the phase mask yielding

$$P(u,v) = [|F_1(u,v)|^2 - |F_2(u,v)|^2] \times \Phi(u,v)$$

$$= 2[\sum_{i=1}^{n} R(u,v)T^*(u,v)\exp[j(vy_0 - vy_i)] \times \Phi(u,v) + \sum_{i=1}^{n} R^*(u,v)T_i(u,v)\exp[-j(vy_0 - vy_i)] \times \Phi^*(u,v)] \times \Phi(u,v) \quad (22)$$

$$= 2\sum_{i=1}^{n} R(u,v)T_i^*(u,v)\exp[j(vy_0 - vy_i)] \times \Phi^2(u,v) + 2\sum_{i=1}^{n} R^*(u,v)T_i(u,v)\exp[-j(vy_0 - vy_i)]$$

It is obvious from Eq. (22) that the zero-order term and the false alarms are eliminated due to the application of the random phase mask. The modified JPS obtained in Eq. (22) is then multiplied by the FAF transfer function, yielding

$$P_{faf}(u,v) = P(u,v) \times H(u,v)$$

$$= \frac{2C(u,v)}{D(u,v) + |R(u,v)|^2} \times [\sum_{i=1}^{n} R(u,v)T_i^*(u,v)\exp[j(vy_0 - vy_i)] \times \Phi^2(u,v) \quad (23)$$

$$+ \sum_{i=1}^{n} R^*(u,v)T_i(u,v)\exp[-j(vy_0 - vy_i)]]$$

Finally, the inverse Fourier transformation of Eq. (23) produces the desired correlation output. Consider that the input scene contains two objects, where $t_1(x,y\text{-}y_1)$ is the target object and $t_2(x,y\text{-}y_2)$ is the non-target object. Assuming $C(u,v)$=1 for unity gain and $|R(u,v)|^2 >> D(u,v)$, Eq. (23) becomes

$$P_{faf}(u,v) \approx 2P(u,v) \times |R(u,v)|^{-2}$$

$$= 2\exp[-(vy_0 - jvy_1)] + 2\exp[j(vy_0 - vy_1)] \times \Phi^2(u,v)$$

$$+ 2\frac{|T_2(u,v)|}{|R(u,v)|}\exp[j(\Phi_r(u,v) - \Phi_{t2}(u,v) + vy_0 - vy_2)] \times \Phi^2(u,v) \quad (24)$$

$$+ 2\frac{|T_2(u,v)|}{|R(u,v)|}\exp[-j(\Phi_r(u,v) - \Phi_{t2}(u,v) + vy_0 - vy_2)]$$

Here, the first term produces a delta function like correlation output for the target object, the second and third terms steer out of the correlation plane due to the random phase mask and the forth term contributes negligible correlation output due to the mismatch between the reference and non-target object.

3 SIMULATION RESULTS

The performance of the proposed SPJTC technique is investigated by considering several gray-level input scenes containing different combinations of a apc image (as target) and a truck image (as non-target) under noise-free and noisy environments as shown in Figures 2 to 5. The apc image in the lower half of the input joint images represents the reference image while various sets of five objects in the upper half represents the unknown input scene. The simulation tests are performed using the MATLAB software package.

To comparae the performance of the PJTC technique with the proposed SPJTC technique, a noise free input scene with five identical targets (five apc images) is considered as shown in Fig. 2(a). The correlation outputs for the PJTC and SPJTC techniques are shown in Figures 2(b) and 2(c), respectively. Here the same filter parameters, $C(u,v) = 1$ and $D(u,v) = 10^{-4}$, are used for both techniques. Figures 2(b) and 2(c) show that both PJTC and SPJTC techniques produce one correlation peak per target, however, the correlation peak intensity (CPI) is higher in SPJTC than that in PJTC. The same simulation test is performed by setting $D(u,v) = 10^{-3}$ and $C(u,v)$=1 and the corresponding correlation outputs for PJTC and SPJTC are shown in Figures 2(d) and 2(e) respectively. Here, both techniques yield almost equal detection performance.

Next the simulation test is performed for another combination of noise free input scene as shown in Figure 3 and the corresponding results are summarized in Tables I and II for $D(u,v)=10^{-4}$ and $D(u,v)=10^{-3}$, respectively. Figure 3 shows the result for one case - the input scene contains three apcs and two trucks. From the results summarized in Tables I and II, we observe that both techniques can successfully detect all targets. However, the SPJTC provides higher CPI than the PJTC. For input scenes containing five targets, the ratio between the maximum correlation peak obtained from SPJTC and that from PJTC is 5.69 whereas the ratio becomes 5.05 when a single target and four non-targets are present in the input scene with $D(u,v)=10^{-4}$. For $D(u,v)=10^{-3}$, the corresponding ratios become 5.58 and 5.1, respectively.

To investigate the noise robustness of the proposed SPJTC technique, Gaussian white noise with signal-to-noise ratio (SNR) of 0 dB was added to the input scene as shown in Figures 4(a) and 5(a), respectively. It is obvious from the correlation outputs shown in Figures 4(b), 4(c), 5(b) and 5(c) that the SPJTC technique yields distinct correlation peaks with the same FAF parameters used in the noise-free case. However, PJTC does not yield correlation output.

To detect targets in noisy input scenes using the PJTC technique, the FAF parameter $D(u,v)$ should be adjusted to a value greater than that used for the noise-free scene. The PJTC yields correlation output for noisy input scene with $D(u,v)=10^{-3}$ only when the number of targets are more than 2 as shown in Figures 4(d) and 5(d). The SPJTC can detect any number of targets either with $D(u,v)=10^{-4}$ or $D(u,v)=10^{-3}$ for both noise free or noisy input scene as shown in Figures 4(c), 4(e), 5(c) and 5(e). Therefore, the SPJTC technique is more robust to noise compared to the PJTC technique. Most importantly, in case of PJTC technique, the FAF parameter needs to be readjusted for noisy input scenes. For various cases of noisy input scenes, the results are summarized in Tables III and IV for $D(u,v)=10^{-4}$ and $D(u,v)=10^{-3}$, respectively. The summarized results shown in Tables I, II, III and IV indicate that the SPJTC technique produces excellent correlation output irrespective of the number of objects and presence of noise in the input scene, which makes it a viable option for efficient multiple target detection.

Fig. 2(a) Gray level joint input scene without noise of 5 apcs as targets

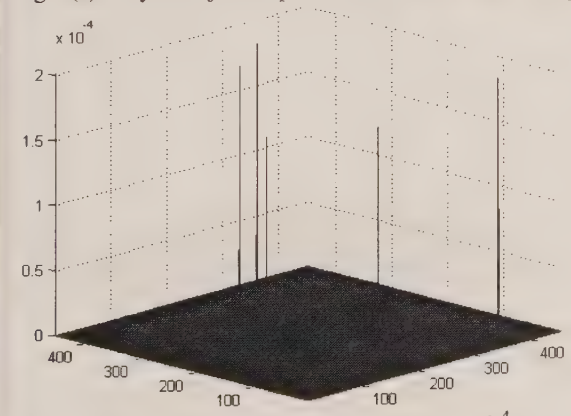


Fig. 2(b) PJTC output of Fig. 2(a) with $D(u,v)=10^{-4}$

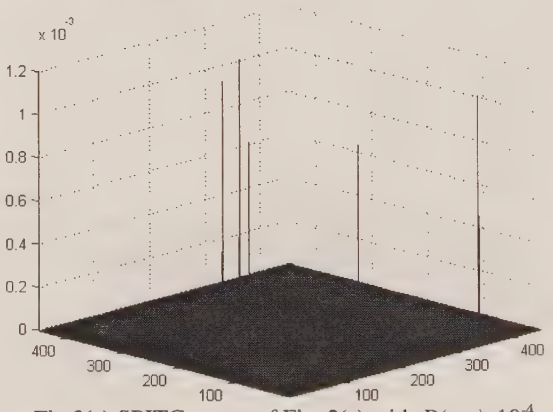


Fig.2(c) SPJTC output of Fig. 2(a) with $D(u,v)=10^{-4}$

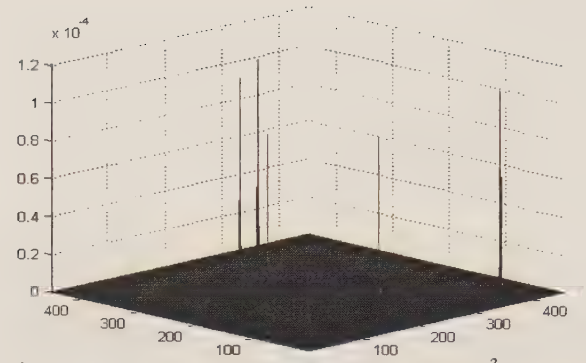


Fig. 2(d) PJTC output of Fig. 2(a) with $D(u,v)=10^{-3}$

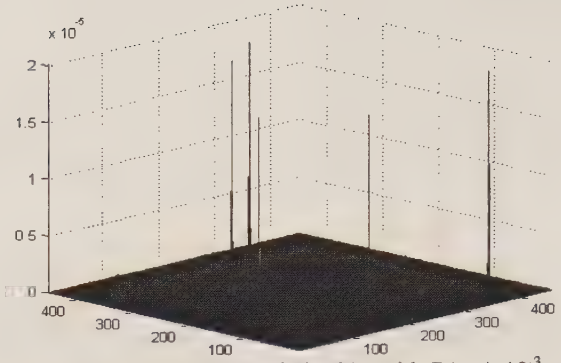


Fig. 2(e) SPJTC output of Fig. 2(a) with $D(u,v)=10^{-3}$

Fig. 3(a) Gray level joint input scene without noise of 3 apcs as targets

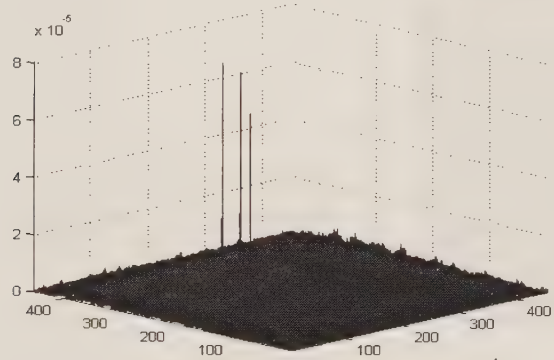


Fig. 3(b) PJTC output of Fig. 3(a) with $D(u,v)=10^{-4}$

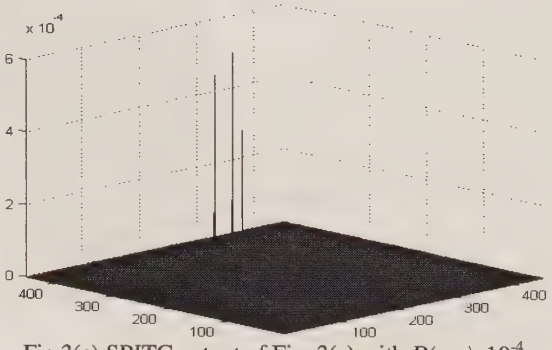


Fig.3(c) SPJTC output of Fig. 3(a) with $D(u,v)=10^{-4}$

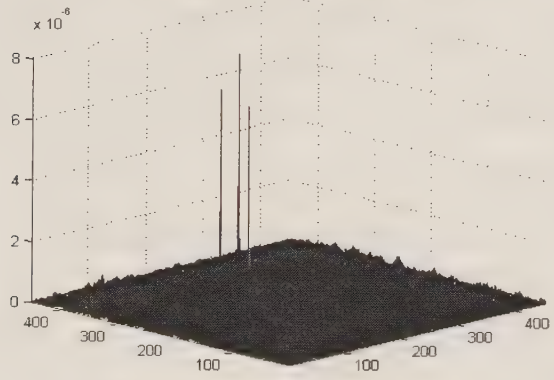


Fig. 3(d) PJTC output of Fig. 3(a) with $D(u,v)=10^{-3}$

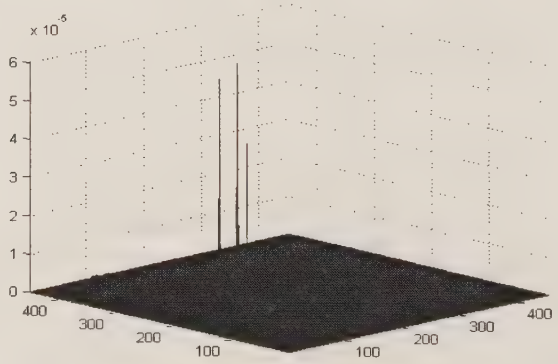


Fig.3(e) SPJTC output of Fig. 3(a) with $D(u,v)=10^{-3}$

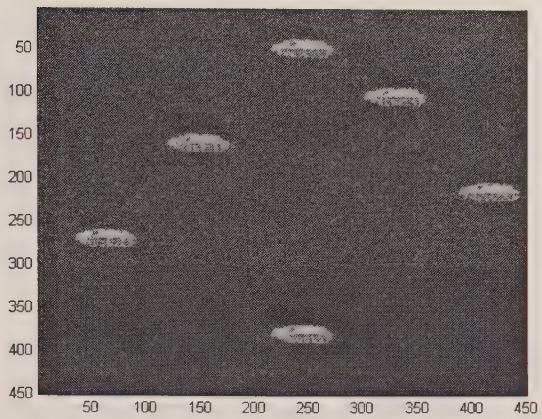


Fig. 4(a) Gray level joint input scene with noise of 5 apcs as targets

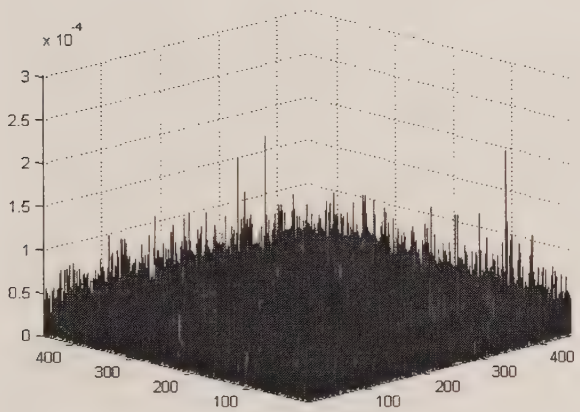


Fig. 4(b) PJTC output of Fig. 4(a) with $D(u,v)=10^{-4}$

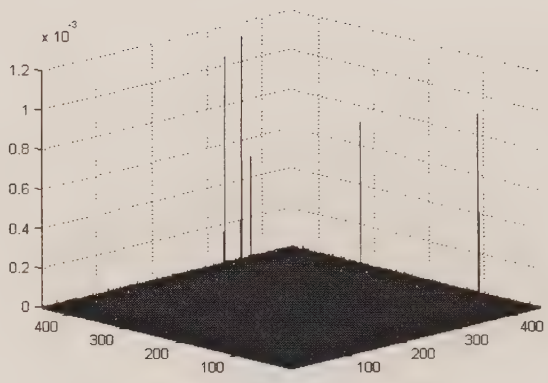


Fig.4(c) SPJTC output of Fig. 4(a) with $D(u,v)=10^{-4}$

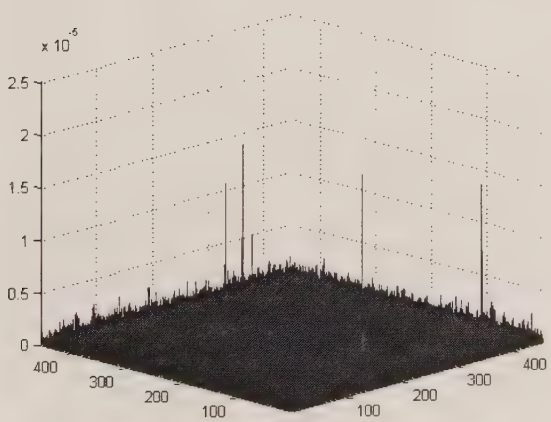


Fig. 4(d) PJTC output of Fig. 4(a) with $D(u,v)=10^{-3}$

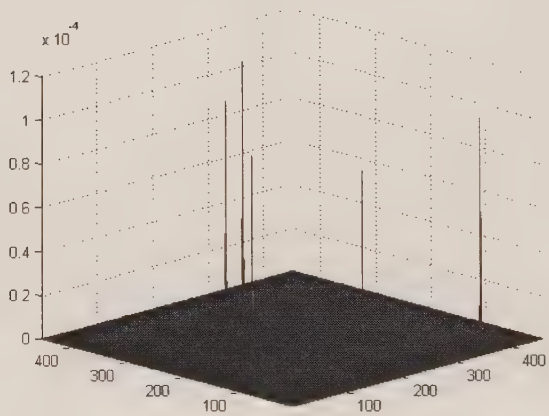


Fig. 4(e) SPJTC output of Fig. 4(a) with $D(u,v)=10^{-3}$

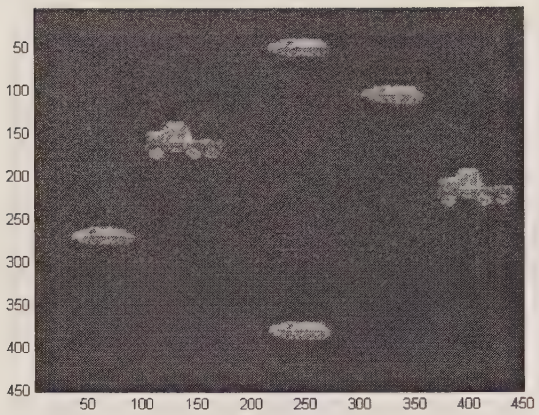


Fig. 5(a) Gray level joint input scene with noise of 3 apcs as targets

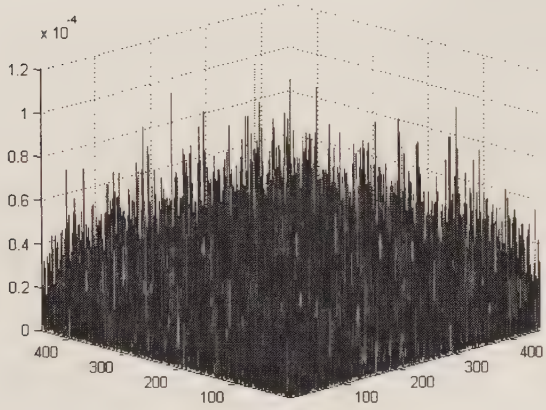


Fig. 5(b) PJTC output of Fig. 5(a) with $D(u,v)=10^{-4}$

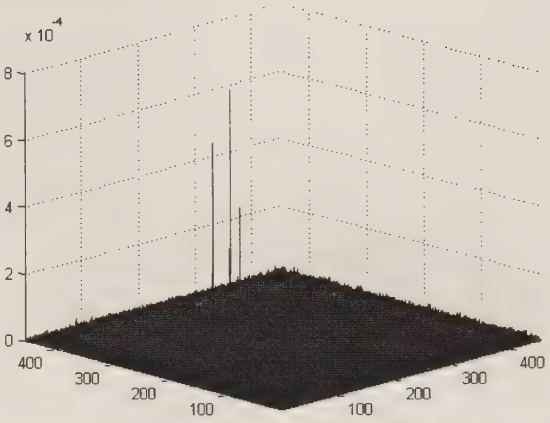


Fig.5(c) SPJTC output of Fig. 5(a) with $D(u,v)=10^{-4}$

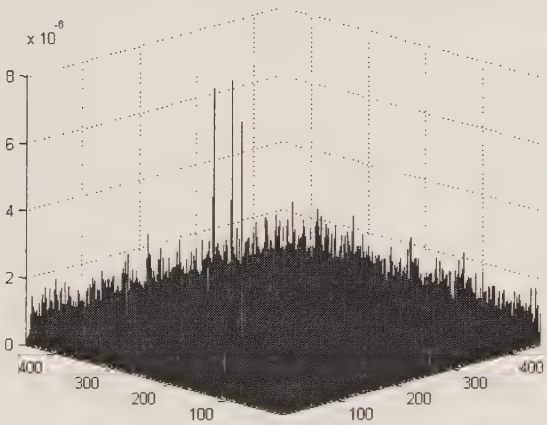


Fig. 5(d) PJTC output of Fig. 5(a) with $D(u,v)=10^{-3}$

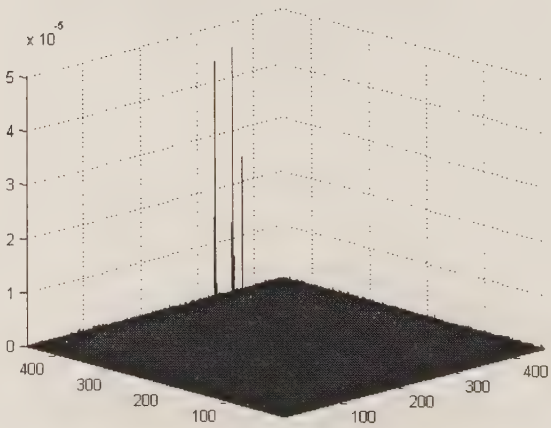


Fig. 5(e) SPJTC output of Fig. 5(a) with $D(u,v)=10^{-3}$

Table I: Performance comparison for noise free input scene with $C(u,v)$=1 & $D(u,v)$=10^{-4}

Input Scene Objects without Noise	PJTC			PSPJTC		
	$D(u,v)$	$Peak_{max}$	$Peak_{min}$	$D(u,v)$	$Peak_{max}$	$Peak_{min}$
5 apcs	10^{-4}	$1.934x10^{-4}$	$1.842x10^{-4}$	10^{-4}	$1.1x10^{-3}$	$1.1x10^{-3}$
4 apcs, 1 truck	10^{-4}	$1.17x10^{-4}$	$1.114x10^{-4}$	10^{-4}	$8.366x10^{-4}$	$8.25x10^{-4}$
3 apcs, 2 trucks	10^{-4}	$7.54x10^{-5}$	$6.45x10^{-5}$	10^{-4}	$5.276x10^{-4}$	$5.05x10^{-4}$
2 apcs, 3 trucks	10^{-4}	$5.94x10^{-5}$	$4.68x10^{-5}$	10^{-4}	$3.34x10^{-4}$	$3.13x10^{-4}$
1 apc, 4 trucks	10^{-4}	$4.2164x10^{-5}$		10^{-4}	$2.1279x10^{-4}$	

Table II: Performance comparison for noise free input scene with $C(u,v)$=1 & $D(u,v)$=10^{-3}

Input Scene Objects without Noise	PJTC			SPJTC		
	$D(u,v)$	$Peak_{max}$	$Peak_{min}$	$D(u,v)$	$Peak_{max}$	$Peak_{min}$
5 apcs	10^{-3}	$1.9x10^{-5}$	$1.86x10^{-5}$	10^{-3}	$1.06x10^{-4}$	$1.035x10^{-4}$
4 apcs, 1 truck	10^{-3}	$1.15x10^{-5}$	$1.025x10^{-5}$	10^{-3}	$8.3x10^{-5}$	$8.04x10^{-5}$
3 apcs, 2 trucks	10^{-3}	$7.76x10^{-6}$	$6.41x10^{-6}$	10^{-3}	$5.15x10^{-5}$	$4.89x10^{-5}$
2 apcs, 3 trucks	10^{-3}	$5.93x10^{-6}$	$5.04x10^{-6}$	10^{-3}	$3.21x10^{-5}$	$3.02x10^{-5}$
1 apc, 4 trucks	10^{-3}	$4.04x10^{-6}$		10^{-3}	$2.061x10^{-5}$	

Table III: Performance comparison for noisy input scene with $C(u,v)$=1 & $D(u,v)$=10^{-4}

Input Scene Objects with Noise	PJTC			SPJTC		
	$D(u,v)$	$Peak_{max}$	$Peak_{min}$	$D(u,v)$	$Peak_{max}$	$Peak_{min}$
5 apcs	10^{-4}	Can not detect		10^{-4}	$1.2x10^{-3}$	$9.637x10^{-4}$
4 apcs, 1 truck	10^{-4}	Can not detect		10^{-4}	$9.2253x10^{-4}$	$8.0252x10^{-4}$
3 apcs, 2 trucks	10^{-4}	Can not detect		10^{-4}	$6.2788x10^{-4}$	$5.3059x10^{-4}$
2 apcs, 3 trucks	10^{-4}	Can not detect		10^{-4}	$4.1391x10^{-4}$	$2.9411x10^{-4}$
1 apc, 4 trucks	10^{-4}	Can not detect		10^{-4}	$1.814x10^{-4}$	

Table IV: Performance comparison for noisy input scene with $C(u,v)$=1 & $D(u,v)$=10^{-3}

Input Scene Objects with Noise	PJTC			SPJTC		
	$D(u,v)$	$Peak_{max}$	$Peak_{min}$	$D(u,v)$	$Peak_{max}$	$Peak_{min}$
5 apcs	10^{-3}	$2.058x10^{-5}$	$1.354x10^{-5}$	10^{-3}	$1.084x10^{-4}$	$9.83x10^{-5}$
4 apcs, 1 truck	10^{-3}	$1.0294x10^{-5}$	$6.7386x10^{-6}$	10^{-3}	$9.4513x10^{-5}$	$7.8494x10^{-5}$
3 apcs, 2 trucks	10^{-3}	$7.9703x10^{-6}$	$6.6395x10^{-6}$	10^{-3}	$4.9287x10^{-5}$	$4.3573x10^{-5}$
2 apcs, 3 trucks	10^{-3}	Can not detect		10^{-3}	$3.642x10^{-5}$	$2.9461x10^{-5}$
1 apc, 4 trucks	10^{-3}	Can not detect		10^{-3}	$2.2985x10^{-5}$	

4 CONCLUSION

A novel technique named SPJTC is proposed for improved pattern recognition that employs phase shifting and phase encoding methods along with fringe adjusted filter for optimization of the correlation performance. The proposed SPJTC technique yields single correlation peak per target and thus ensures the best utilization of input and output plane SBP. This technique overcomes the limitations of currently available PJTC technique, which has been found to be sensitive to the number of targets and the presence of noise in the input scene. The proposed SPJTC technique can detect targets from an input scene irrespective of the number of targets or non-targets in noise-free and noisy input scenes with the same system parameters. Simulation confirms the effectiveness of the proposed technique for multiple target detection.

ACKNOWLEDGEMENT

This research was supported by a grant from the Army Research Office (ARO: 43004-CI). The authors wish to thank Richard Sims and William Sanders for many rewarding discussions.

REFERENCES

1. A. B. VanderLugt, "Signal detection by complex spatial filtering," IEEE Trans. Inf. Theory, **IT-10**, 139- 145 (1964).
2. C. S. Weaver and J. W. Goodman, "Technique for optically convolving two functions," Appl. Opt., **5**, 1248 - 1249 (1966).
3. J. C. Kirsch and D. A. Gregory, "Video rate optical correlation using a magneto-optic spatial light modulator," Opt. Eng., **29**, 1122-1128 (1990).
4. F. T. S. Yu and X. J. Lu, "A real time programmable joint transform correlator," Opt. Commun., **52**, 10 (1984).
5. B. Javidi and C. Kuo, "Joint transform image correlation using a binary spatial light modulator at the Fourier plane," Appl. Opt., **27**, 663 (1988).
6. F. T. S. Yu, C. Li and S. Yin, "Comparision of detection efficiency for nonzero order and conventional joint transform correlation," Opt. Eng., **37**(1), 52-57 (1998).
7. Q. Tang and B. Javidi, "Multiple-object detection with a chirp-encoded joint transform correlator," Appl. Opt., **32**, 4344 (1993).
8. F. T. J. Johnson, T. H. Barnes, T. Eiju, T. G. Haskell and K. Matsuda, "Analysis of a joint transform correlator using a phase-only spatial light modulator," Opt. Eng., **30**(12), 1947-1957 (1991).
9. W. B. Hahn and D. L. Flannery, "Design elements of a binary joint transform correlator and selected optimization techniques," Opt. Eng., **31**(5), 896-905 (1992).
10. F. T. S. Yu, F. Cheng, T. Nagata, and D. A. Gregory, "Effects of fringe binarization of multi-object joint transform correlation," Opt. Eng., **28**, 2988 (1989).
11. G. Lu. Z. Zhang, S. Wu, and F. T. S. Yu, "Implementation of a non-zero-order joint-transform correlator by use of phase-shifting techniques." Appl. Opt., **36**, 470-483 (1997).
12. H. J. Su and M. A. Karim, "Performance improvement of a phase-shifting joint transform correlator by use of phase-iterative techniques." Appl. Opt., **37**, 3639-3642 (1998).
13. M. S. Alam and M. A. Karim, "Fringe-adjusted joint transform correlator," Appl. Opt., **32**, 4344 (1993).
14. M. S. Alam and M. A. Karim, "Joint transform correlation under varying illumination," Appl. opt., **32**, 4344 (1993).
15. C. T. Li, S. Yin, and F. T. S. Yu, "Nonzero-order joint transform correlator," Opt. Eng., **37**(1), 58-65 (1998).
16. M. S. Alam and D. Chain, "Efficient multiple target detection using a joint wavelet transform processor," Opt. Eng., **39**(5), 1203-1210 (2000).
17. C. Huaixin, C. Zhenpei, and C. Xiaowei, "Binary joint transform correlator subtracting a power spectrum based on phase-shift technique," Chin. J. Lasers., **25**(4), 353-358, (1998).
18. M. S. Alam, "Phase-encoded fringe-adjusted joint transform correlation," Opt. Eng., **39**(5), 1169-1176 (2000).

On the Development of Filter Management Module for Grayscale Optical Correlator

Hanying Zhou, Casey Hughlett,* Jay C. Hanan, and Tien-hsin Chao

Jet Propulsion Laboratory, MS 303-300
California Institute of Technology
4800 Oak Grove Drive, Pasadena, CA 91109-8099

*Zlabs, 900 Sandhurst Dr. Edmond, OK 73034

ABSTRACT

An Optical Processing for the Mining and Identification of Targets (OPMIT) system is being proposed to significantly reduce broad area search workload for NIMA imagery analysts. Central to the system is a Grayscale Optical Correlator (GOC), developed by JPL in recent years. In this paper we discuss some preliminary development of an important system component – the filter management module – that is critical for the success of GOC operation. The emphasis is on the streamlining the OT-MACH filter synthesis/testing procedure for effective and efficient filter design while maintaining filter performance.

Keywords: Filter synthesis, gray-scale optical correlator

1. INTRODUCTION

Many government agencies rely on imagery analysis to gain valuable intelligence. As the volume of imagery available from various sensors and platforms continually increase, the demands on imagery analysts also increase dramatically, from the static conditions during the Cold War to the more dynamic, continually spreading and evolving threat of global terrorism, from primarily strategic analysis to the need for more real-time support to operations. Due to the cost and time lag in obtaining fully qualified image analysts, the intelligence community desires new technology tools to significantly reduce the analyst's broad area search workload, freeing them to work on more in-depth and complex analytical problems and on areas of high probably for target location

JPL, under years of sponsorship from DOD and NASA, has developed a Grayscale Optical Correlator (GOC) for high speed ATR application.[1-3] The GOC takes advantages of shift-invariant parallel Fourier transform achieved by optical lenses at the speed of light and with minimal electrical power consumption. In operation, tens or hundreds templates (pre-computed "composite filters" in the Fourier domain) are uploaded and checked against each incoming input frame to detect and locate the target-of-interest in the input scene (with sharp correlation peak in the output correlation plane). Due to the inherent shift-invariant characteristics of the optical correlator, target(s) located in anywhere in the input plane will be simultaneously detected over the location(s) of the corresponding target(s).

The interest in optical processing for the ATR applications stems from the fact that optical processing inherently offers massive parallelism, ultrahigh speed, and shift invariance, all with large frame size, not easily achievable by electronics counterpart. The optical processing speed of GOC is mainly limited by the updating speed of the input or the filter devices. Currently this speed is 1000 frames/sec with a 512x512 frame size, and will remain roughly the same with even larger frame size. On the contrary, a digital processing speed will decrease proportional to $(1/N^2)$ where N is the pixel number of the input in each of the two dimensions. Compared to the state-of-the-art digital processor (the Motorola G4 processor using a MPC7445 CPU), the GOC processing speed is about 1000 times faster.

While the speed enhancement in using optical over digital processes is significant, using optical correlator alone may not have the necessary processing accuracy as digital tools have. To optimally utilize optical correlator for

Optical Pattern Recognition XV, edited by David P. Casasent, Tien-Hsin Chao, Proc. of SPIE Vol. 5437 (SPIE, Bellingham, WA, 2004) · 0277-786X/04/$15 · doi: 10.1117/12.548066

ATR, additional digital tools, algorithms and approaches need to be integrated to improve the processing accuracy and reduce false positive/negative alerts while retain the speed enhancements. This is the idea behind the Optical Processing for the Mining and Identification of Targets (OPMIT) system that is recently envisioned at JPL for NIMA application. By combining the broad-area search capability with rapid and reliable ATR, the OPMIT will enhance the Tasking-Processing-Exploitation-Dissemination process by rapidly identifying prospective targets and generating the necessary information flows for timely response to target opportunities.

2. OVERVIEW OF THE OPMIT SYSTEM

The purpose of the OPMIT system is to reduce the analyst's workload to free them from more routine search activities to concentrate their time on more focused and complex analyses. As a result, the primary mode of operation is to process vast amount of incoming and on-line imagery during the upstream processing stages based on user defined search requests and then to provide the search results to the analyst for analysis.

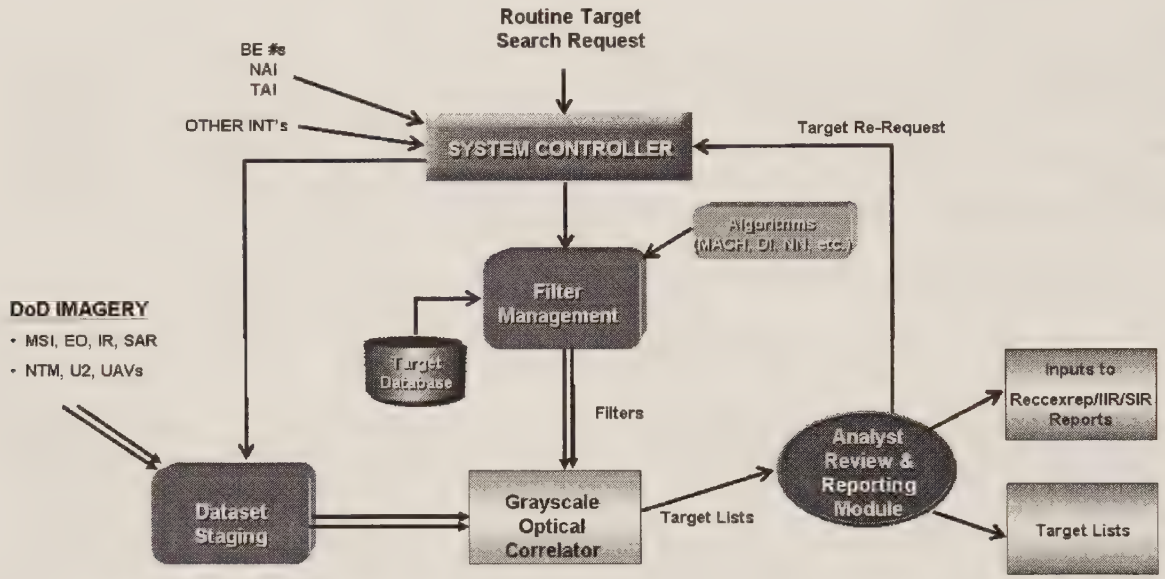


Fig.1 OPMIT system blocks

The initial main system components (see Fig.1) consist mostly of the capabilities under development at JPL for or have already been delivered to the NIMA. Central to the OPMIT system is the **Grayscale Optical Correlator (GOC)**, which is a hardware/firmware subsystem that has the ability to accept and analyze imagery segments and compare those segments against a target "filter" containing the desired target. A **Dataset Staging** module acquires datasets for rapid loading into the GOC. The GOC uses filters when searching for targets in a dataset and filters must be pre-configured with examples of the search target. A **Filter Management** module manages the entire filter process from filter pre-processing (target retrieval from the Target Database, algorithm execution and filter storage), filter retrieval from storage and filter loading into the GOC. Initially, the algorithm used to create filters will be based on the modified Optimal-Tradeoff Maximum Average Correlation Height (OT-MACH) algorithm,[4-6] but future deliveries will allow (and even encourage) other algorithm use for maximum search accuracy. A **Target Database** module provides a repository for target images and characteristics in a standardized structure. The template-matching based ATR approach requires the acquisition of a large sample set of images that represents the variety of azimuth angles, elevation angles, scales, illumination conditions, and in-plane rotation for each target of interest. All forms of sample and example target views and descriptions will be captured in the Target Database for use by the system. It will be initially populated by available electronic sources and augmented by the creation of additional spectral, perspective, and lighting/obscured target sets through manual means (e.g. model re-creations) to

assure satisfactory identification rates. During later deliveries, the Target Database will be populated by more automated means via a Target Build-Up Workstation. An **Analyst Review and Reporting** module will manage the raw results received from the GOC and convert them into meaningful, easy-to-use target representation for use by the analyst and for inclusion in standard NIMA reports. Target results are matched with user requests. It will contain a Graphical User Interface (GUI) allowing direct user interaction with the target results and an standard interface will be provided for linkage to other existing NIMA applications for the exchange of target search results. Controlling system operation will be the responsibility of the **System Controller** module which has overall oversight of system operation, specifying the datasets to be loaded for search, coordinating user requests with the search results and providing the user interface for both system administration and for user target search requests.

3. FILTER MANAGEMENT MODULE

One of the important components of the proposed OPMIT system is the Filter Management (FM) module. The GOC uses filters when searching for targets in a dataset and filters must be pre-synthesized with examples of the search target. The FM module manages the entire filter process (see Fig.2) from target selection and retrieval from the target database, filter synthesis through selected algorithm, filter storage, filter selection and retrieval from storage and filter loading into the GOC.

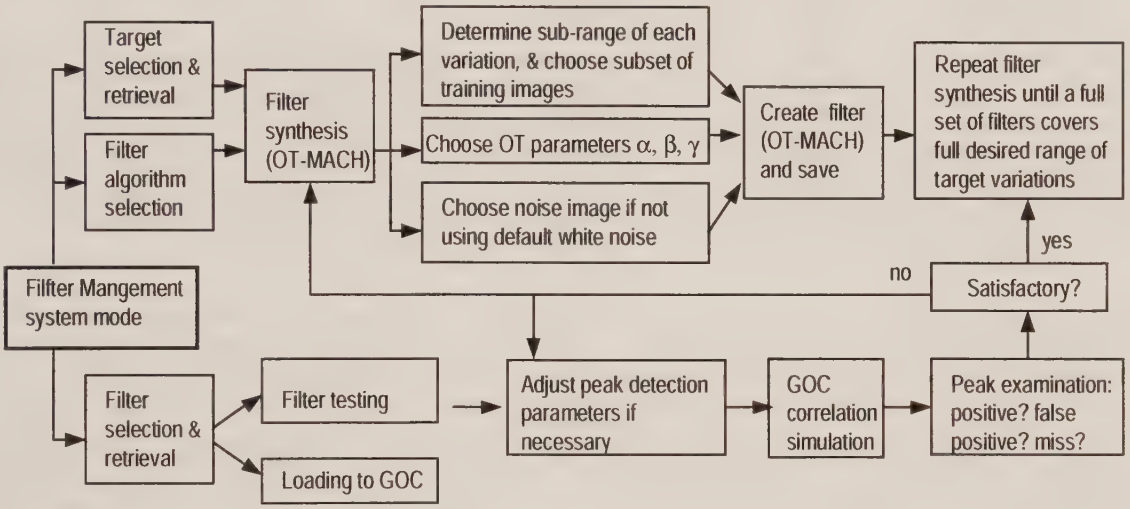

Target selection & retrieval
Filter algorithm selection
Filter synthesis (OT-MACH)
Determine sub-range of each variation, & choose subset of training images
Choose OT parameters α, β, γ
Choose noise image if not using default white noise
Create filter (OT-MACH) and save
Repeat filter synthesis until a full set of filters covers full desired range of target variations
yes
no
Satisfactory?
Filfter Mangement system mode
Filter selection & retrieval
Filter testing
Loading to GOC
Adjust peak detection parameters if necessary
GOC correlation simulation
Peak examination: positive? false positive? miss?

Fig.2 Filter Management Module blocks

3.1 OT-MACH filter algorithm

Though other FFT-based algorithms will be (developed and) evaluated for using with GOC, initially we will use OT-MACH (optimum trade-off Maximum Average Correlation Height) based filters. OT-MACH filter is among the few algorithms that are formulated with optimal implementation on optical correlator in mind. It has the best chance to deliver superb performance on GOC.

The OT-MACH filter tries to balance simultaneously several conflicting performance measures such as Average Correlation Height (ACH), Average Similarity Measure (ASM), Average Correlation Energy (ACE) and Output Noise Variance (ONV), by minimizing the following energy function:

$$E(\mathbf{h}) = \alpha(ONV) + \beta(ACE) + \gamma(ASM) - \delta(ACH)$$
$$= \alpha\mathbf{h}^{+}\mathbf{C}\mathbf{h} + \beta\mathbf{h}^{+}\mathbf{D}_x\mathbf{h} + \gamma\mathbf{h}^{+}\mathbf{S}_x\mathbf{h} - \delta\left|\mathbf{h}^{T}m_x\right|$$

The resulting optimal-tradeoff MACH filter (in frequency domain) is given by [4-5]

$$\mathbf{h} = \frac{\mathbf{m}_x^{*}}{\alpha\mathbf{C} + \beta\mathbf{D}_x + \gamma\mathbf{S}_x}$$

where α, β and γ are nonnegative OT parameters. m_x is the average of the training image vectors x_1, x_2, ...x_N. C and D_x are the diagonal power spectral density matrixes of additive input noise and of the training images, respectively. S_x is the similarity matrix of the training images: $S_x=\Sigma(X_i-M_x)^{*}(X_i-M_x)/N$, with M_x denoting the average of X_i.

By choosing different values of α, β and γ, one can control the OT-MACH filter's behavior to suit different application requirements. For example, when $\beta=\gamma=0$, the resulting filter behaves much like a MVSDF filter with relative good noise tolerance but broad peaks. If $\alpha=\gamma=0$, then the filter is more like a MACE filter, which generally gives sharp peaks and good clutter suppression but is very sensitive to distortion. For $\alpha=\beta=0$, the filter is a MACH filter which itself is designed with high tolerance for distortion.

Often in many ATR applications a potential target is embedded in a cluttered background. We have recently further improved the performance of OT-MACH filters in a severe cluttered/noisy background.[6] The idea is to have better representation of input noise characteristics in the filter design even though the exact knowledge about the input noise may not be available. Particularly we have demonstrated that it is possible to use a simulated noise/clutter image to substitute the often over simple white noise covariance in the filter design. This of course assumes that one would have some information about the typical or the most likely noise/clutter environment that poses the main difficulty in target detection.

To implement the OT-MACH filter on optical correlator, one needs to find out its minimum Euclidean distance (MED) version of the filter for the SLM coding domain.[7] To better utilize the limited dynamic range of the currently available SLMs, we developed a practical filter dynamic range compression technique, the histogram stretch, after using the MED projection.[8] The procedure for this histogram stretch is rather simple. First, we calculate the standard deviation, σ, of the pixel histogram of the MED version filter. The smaller the standard deviation σ, the narrower the filter's histogram distribution. Then we find out pixels whose values are outside a certain range of pixel values, say, +/-3σ, and assign those pixels with maximum values (in our 8-bit real-valued SLM case, it is +/-127) and linearly stretch the rest pixel values toward the maximum values.

Variations of OT-MACH algorithms will be evaluated and provided, if necessary, as potential alternatives or compliments to the standard version of OT-MACH algorithm. Other GOTS and COTS FFT based algorithms will also be evaluated. We will also develop optimization procedure that will be necessary for those filters not designed for optical implementation.

3.2 Generating Filter Bank with a Streamlined Procedure

Using above OT-MACH filter algorithm one can generate distortion tolerant filters as needed. However, this can often be a frustrating and time-consuming task even for an experienced filter designer. Although FFT-based filters are shift invariant, they are nevertheless sensitive to scale, rotation, perspective, and lighting conditions. In addition, targets may also be partially occluded, camouflaged or may be decoys. In order to achieve the highest level of assurance that the targets are identified with minimal errors while the number of filters is reasonably small, all desired variations and distortions in targets need to be carefully included systematically when generating filters. That is, one needs to generate a filter bank for the target in which each filter covers a small range of variations, and a set of filters covers the full expected range of those variations.

But how much each variations should be included in a single filter is very much application scenario dependent, and can be found generally only by try and error. To facilitate this process, we are currently developing a software package to streamline the filter synthesis/testing procedure effectively while maintain OT-MACH filter performance. The flowchart for this procedure is outlined in Figures 2 and 3. Initially, a small subset of training images that consist of a sub-range of size, orientation, aspect, etc., variations are chosen from the target dataset. The ranges of these variations will be set according to experience or from default, e.g., set the size variation in a single filter to 20% (with 5 intervals, 0, +/-10%, and +/-20%), orientation variation to 40° (with 5 intervals, 0, +/-10°, and +/-20°), perspective variations to 30° (with 5 intervals, 0, +/-10°, and +/-20°). Next, choose optimal OT parameters according to experience or start from default, e.g., set α=0.3, β=0.3 and γ=1 to emphasize good distortion tolerance. Initially the default white noise can be used. With these parameters chosen one can synthesis filter.

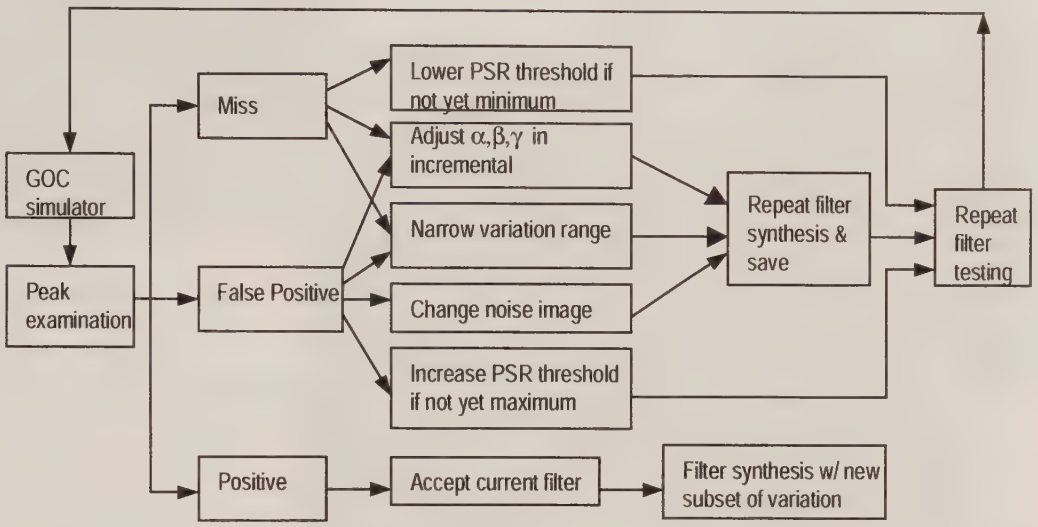


Fig.3 Filter test and re-synthesis flow chart

The filter designer (the intelligence analyzer) will now be able to go through the build-in GOC simulator with representative testing images to examine the correlation outcomes. All the desired peak locations/areas will be specified through a GUI for each testing image in advance. If desired peaks (positive) occur with better than pre-specified PSR(s) and peak intensities(s) and with no false positive, the filter will be accepted without modification. If desired peak is missing, PSR threshold will be lowered first (unless it already reaches a pre-specified minimum, e.g., 10). If this does not help, OT parameters α, β, and γ will be then adjusted (e.g., to [1 0.3 0.3] for good noise tolerance or to [0.3 1 0.3] for sharp peak first, and afterward in a pre-specified incremental, e.g., 0.2) until the best combination of parameter choices are found. Experienced filter designer can also input parameters α, β, and γ directly or narrower their search ranges (e.g., for the purpose of maintaining sufficient optical efficiency of filters, etc.). After exhausted on this, one can improve filter performance by narrowing variation ranges, e.g., limit the size variation in a single filter to 10% (with 3 intervals, 0 and +/-10%), orientation variation to 20° (with 3 intervals, 0 and +/-10°), perspective variations to 10° (with 3 intervals, 0, +/-10°). Once the appropriate variation subranges are found, one can fine tune parameters α, β, and γ within a smaller range of the previous found optimal combination. After several iterations, it should usually result in a good OT-MACH filter covering sub-ranges of desired variations. Then one can move on to synthesize filters that will cover the rest of the variation range. The parameters and variation range for the rest filters will be easily found out since they usually change little for target of similar characteristics.

3.3 An Example

One of our recent efforts involves filter synthesis for landing site tracking with GOC for NASA application. The focus here is on synthesizing filters systematically with multiple types of variation tolerance, which includes

tolerances of trajectory change, lighting difference, spinning motion, and altitude change. These are important under real mission conditions.

As depicted in Figure 4, the landing site (a MER '03 candidate site, Eos Chasma VM41A2) is in the floor of a valley. The simulation scene starts at about 55km. The spacecraft quickly approaches the pre-selected landing site with fast spinning and a downward speed about 7km/s. The target here is a crater, which is hardly seen at the beginning but is quite clear at the later stage. The change in size of the target is about 4:1. Due to the spinning motion of spacecraft, the orientation of the target can be any way in the input frame (rotation change, 360°). Another important and realistic variation of the target comes from the trajectory change. In our current effort we included 10° variation in trajectory change (perspective change, 10°). In addition, we also included lighting effect (texture change resulting from 90° "sun" illumination change, see Fig.5).

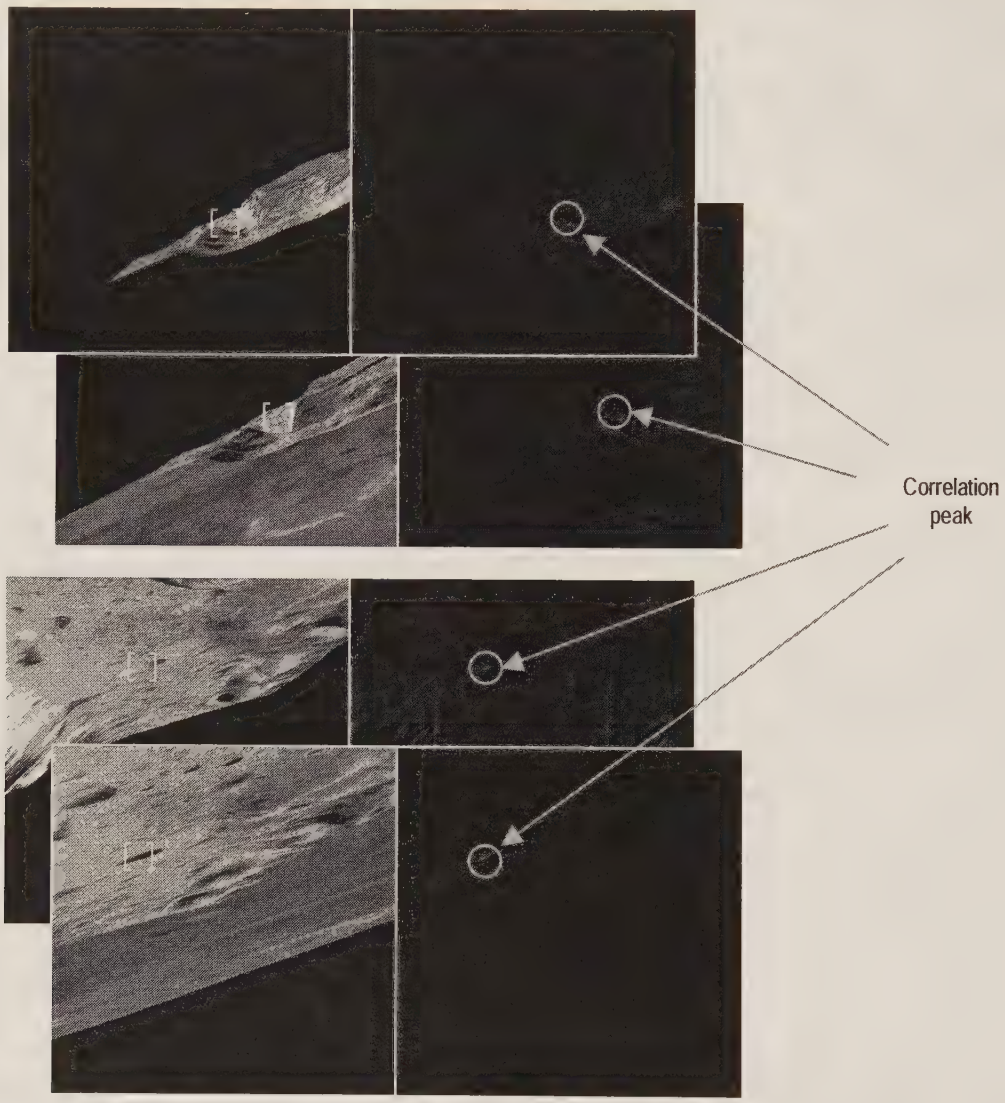


Fig.4 In addition to significant scale, orientation, and lighting variation tolerances, the filters are also tolerant of up to 10° variation in trajectory (perspective change)

In Fig.4, each pair of images frame use the same filter, while each different stage uses a different filter. A total of 25 filters were generated to cover those range of variations; each filter covers sub ranges of those variations (36° rotation,+/-10~20% in size, 10° view, 90° sun direction). Compared with a previous similar simulation but with no

perspective and texture changes included,[9] the number of filters needed increases about 3 fold. Apparently, the more variations one wish to include (e.g., perspective changes in other directions), the more filters will be needed.

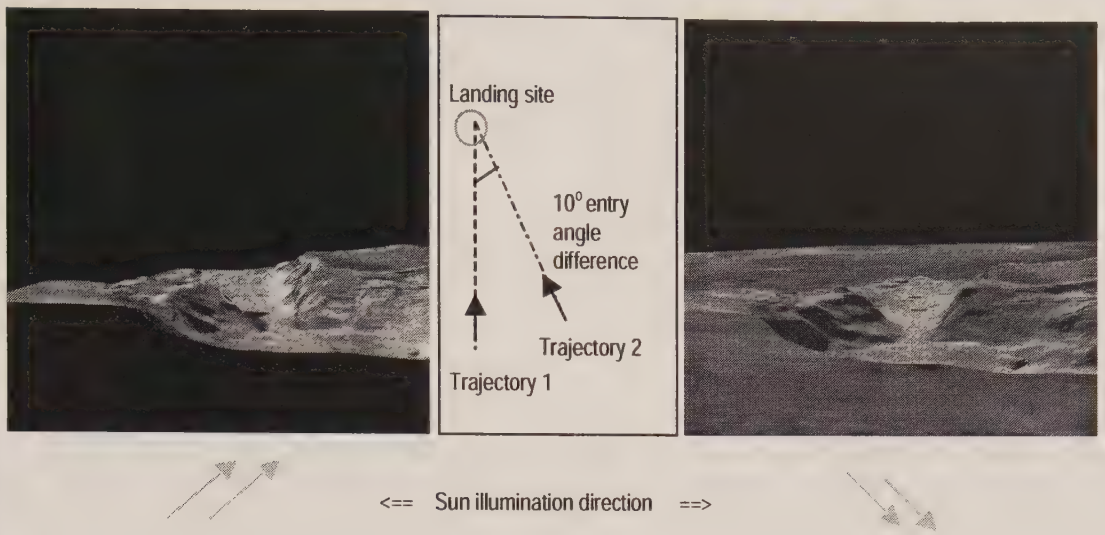


Fig. 5 Illustration of trajectories and lighting illumination directions

4. CONCLUSION

We have described our preliminary effort on the development of Filter Management module, which is part of the GOC-based OPMIT system recently proposed for NIMA image analysis application. The main effort was focused on the streamlined filter synthesis and testing procedure to create needed filter bank efficiently and effectively while maintain filter performance. An example is given. Further development is underway.

ACKNOWLEDGMENTS

The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

REFERENCES

1. Tien-Hsin Chao, George Reyes, and Youngchul Park, "Grayscale Optical Correlator", SPIE v.3386, 1998.
2. Tien-Hsin Chao, George Reyes, and Hanying Zhou, "Automatic Target Recognition Field Demo Using a Grayscale Optical Correlator", SPIE v.3715, p.399-406, 1999
3. Tien-Hsin Chao, Hanying Zhou, and George Reyes, "Compact 512x512 Grayscale Optical Correlator", SPIE v.4734, p.9-12, 2002.
4. B. V. K. Vijaya Kumar, D. Carlson, and A. Mahalanobis, "Optimal trade-off synthetic discriminant function filters for arbitrary devices," *Opt. Lett.* 19, pp.1556-1558, 1994.
5. V. Laude and Ph. Refregier, "Multicriteria characterization of coding domains with optimal Fourier spatial light modulator filters," Appl. Opt. 33, pp. 4465-4471, 1994.
6. Hanying Zhou and Tien-Hsin Chao, "MACH Filter Synthesizing for Detecting Targets in Cluttered Environment for Gray-scale Optical Correlator," SPIE **3715**, pp. 394-398, 1999.
7. R. D. Juday, "Optimal realizable filters and the minimum Euclidean distance principle," Appl. Opt. 32, pp.5100-5111, 1993.

8. Hanying Zhou, Tien-Hsin Chao, and George Reyes, "Practical saturated filter for grayscale optical correlator using bipolar-amplitude SLM," SPIE **4043**, 2000.
9. Hanying Zhou, Tien-Hsin Chao, Bryan Martin, and Nate Villaume, "Simulation of miniature optical correlator for future generation of spacecraft precision landing," SPIE **5106**, p.179-185, 2003.

Performance Evaluation of Quadratic Correlation Filters for Target Detection and Discrimination in Infrared Imagery

S. Richard F. Sims
US Army RD&E Command
Aviation and Missile RDEC
Redstone Arsenal, AL.

Abhijit Mahalanobis
Lockheed Martin, MFC, Orlando, FL.

Abstract
The detection and discrimination of targets in infrared imagery has been a challenging problem due to the variability of the target and clutter (background) signatures. In this paper we discuss the application of a novel quadratic filtering method using missile seeker infrared closing sequences. Image filtering techniques are well suited for target detection applications since they avoid the disadvantages of typical pixel-based detection schemes (such as segmentation and edge extraction). Another advantage is that the throughput complexity of the filtering approach, in the detection process, also does not vary with scene content. The performance of the proposed approach is assessed on several data sets, and the results are compared with that of previous linear filtering techniques. Since we can obtain the signature of some of the clutter "in-the-field" or during operation, we examine the impact of updating the filters to adapt to the clutter.

Introduction
Over the past three decades, considerable work has been done in the area of detection and recognition of targets in infra-red (IR) imagery. Popular methods for achieving this goal include statistical techniques based on feature classification [1,2], model based techniques [3,4], and neural-networks [5,6,7]. Generally speaking, the potential targets are segmented from the background; boundary edges and internal features are measured, and fitted to either statistical or structural models. These methods work well when there is sufficient number of pixels on the targets and the contrast boundaries are well defined. In general however, the performance of most segmentation based methods suffers considerably when the targets exhibit poor contrast relative to the background, or when parts of the target are either obscured or in close proximity to clutter objects. In such cases segmentation is a difficult task which generally induces significant error.
In this paper, we focus on target detection techniques based on signal processing methods that do not require segmentation. The leading approach is based on the use of *Correlation Filters* [8,9]. Perhaps the earliest and most well known among these is the matched spatial filter (MSF) [10]. Although the MSF is optimum for the detection of a single (deterministic) pattern in the presence of additive white Gaussian noise (AWGN), it is hardly suitable for robust detection and recognition of targets in real-world images. For our case however, to add robustness and distortion tolerance to the detection process, we consider the data and the pattern as both being a stochastic process. This no longer fits

Optical Pattern Recognition XV, edited by David P. Casasent, Tien-Hsin Chao, Proc. of SPIE Vol. 5437 (SPIE, Bellingham, WA, 2004) · 0277-786X/04/$15 · doi: 10.1117/12.548067

the model in which the MSF is optimal and a different methodology is needed to define correlators for the detection problem. As a result, many approaches have been proposed for the design of filters for correlators that exhibit robust behavior in the presence of clutter and wide variations in the targets' signatures.

As outlined in [8], much work has been done on the design of linear filters for pattern recognition. In addition to locating the target in the input signal (or image), the filter design must accommodate the suppression of side-lobes on the target of interest and false alarms. In this paper, we evaluate a new type of *quadratic Correlation Filter* (QCFs) that improves on linear correlation filters, but retains the advantage of shift-invariance without requiring segmentation or registrations. The method reviewed here for the design of QCFs optimizes a performance metric formulated as a Rayleigh quotient leading to eigen-space solutions for QCFs.

Our focus is on comparing the performance of the maximum average correlation height (MACH) linear filters [11] and the QCFs in the detection of targets and in the subsequent discrimination based on the detections in infrared image "closing" sequences. We plot the detection metrics on the target and other objects versus range. The detection metrics used are a confidence measure for the QCF and the peak-to-side lobe ratio for the MACH. This type of plot or strip chart clearly shows the separation and in some cases non-separation of the target versus "clutter." In our closing sequence scenarios, which are from long range to close range, which varies the resolution on the target, clutter is image dependent such that anything other than the object of interest encoded into the detection filter is deemed to be clutter. This clutter can include other targets, as shown in Fig. 2, even of the same type if they are present in the same image. In addition, the clutter can also clearly include any scene artifacts, man made or natural. As the infrared camera moves from long range to close range, the aspect and perspective of the field of view onto the background changes, along with the range to target. In general, this scene background change and variation in pixel resolution on the target has the potential to drastically vary the target like or clutter objects in the scene.

In the next sections we discuss the theory and implementation of QCF filters, and specifically the version which has been called the Tuned Basis Function (TBF). We then describe the detection metrics, the data we used, and the performance quantification of the TBF versus the MACH filters.

Review of QCF Theory

Although the theory of linear correlation filters (including the MACH filter) are well known, there is less literature available on the design of QCFs. For ease of reference, we now review a technique for the design of QCFs. The issue of finding targets in background clutter, in this case, is a two-class pattern recognition problem. Consider an input training signal $\mathbf{x} = [x(0) \quad x(1) \quad \cdots \quad x(N-1)]^T$ that can be either a target (class ω_1) or background (class ω_2). For now we assume that $\mathbf{x}$ is a purely real signal. The output of the quadratic filter is defined as

$$\mathbf{x}^T\mathbf{T}\mathbf{x} = \sum_{i=1}^{N}\sum_{j=1}^{N} t_{ij}x(i)x(j) \qquad \text{Eq 1}$$

The coefficient matrix $\mathbf{T} = \{t_{ij}\}$ is square and real but otherwise unrestricted. It should be noted that since $\mathbf{T}$ is not restricted to be positive definite, the output can be either positive or negative. The idea is to determine $\mathbf{T}$ such that the output is positive and as large as possible when $\mathbf{x} \in \omega_1$, and negative or as small as possible when $\mathbf{x} \in \omega_2$. Assume that the form of the coefficient matrix for a quadratic filter is

$$\mathbf{T} = \sum_{i=1}^{N} \mathbf{q}_i\mathbf{q}_i^T \qquad \text{Eq 2}$$

where $\mathbf{q}_i$ form an ortho-normal basis set. The output φ of the QCF in response to an input training vector $\mathbf{x}$ is given by

$$\varphi = \mathbf{x}^T\mathbf{T}\mathbf{x} \qquad \text{Eq 3}$$

The idea is to maximize the separation between the expected values of φ for the two classes while reducing an upper bound on the variance of φ. It can be shown that the separation between the expected value for the two classes is given by Eq. 4 where

$$E_1\{\varphi\} - E_2\{\varphi\} = \sum_{i=1}^{N} \mathbf{q}_i(\mathbf{R}_1 - \mathbf{R}_2)\mathbf{q}_i^T \qquad \text{Eq 4}$$

$\mathbf{R}_i$ is the correlation matrix of class ω_i. The maximization of this term alone however results in a solution which does not adequately control the spread of the class statistic. To avoid this, we also seek to minimize

$$E_1\{\varphi\} + E_2\{\varphi\} = \sum_{i=1}^{N} \mathbf{q}_i(\mathbf{R}_1 + \mathbf{R}_2)\mathbf{q}_i^T \qquad \text{Eq 5}$$

It can be shown that including this term places bounds on the second order statistics of φ, and helps to reduce the spread. To achieve our objectives, we seek $\mathbf{q}_i$ to maximize the ratio shown in Eq. 6.

$$J(\mathbf{q}) = \frac{E_1\{\varphi\} - E_2\{\varphi\}}{E_1\{\varphi\} + E_2\{\varphi\}} = \frac{\sum_{i=1}^{N} \mathbf{q}_i(\mathbf{R}_1 - \mathbf{R}_2)\mathbf{q}_i^T}{\sum_{i=1}^{N} \mathbf{q}_i(\mathbf{R}_1 + \mathbf{R}_2)\mathbf{q}_i^T} \qquad \text{Eq 6}$$

By taking the derivative of the expression in Eq. 6 with respect to $\mathbf{q}_i$ and setting it equal to zero, we then with some manipulations have Eq. 7.

$$(\mathbf{R}_1 + \mathbf{R}_2)^{-1}(\mathbf{R}_1 - \mathbf{R}_2)\mathbf{q}_i = \lambda_i\mathbf{q}_i \qquad \text{Eq 7}$$

Thus $\mathbf{q}_i$ is an eigenvector of the matrix $(\mathbf{R}_1+\mathbf{R}_2)^{-1}(\mathbf{R}_1-\mathbf{R}_2)$ with eigenvalue λ_i. The dominant eigenvector of course maximizes the ratio for the choice of any one vector. However by imposing the constraint that $\mathbf{q}_i$ must form an orthonormal set it is seen that the optimum choice for N vectors is to select N eigenvectors of $(\mathbf{R}_1+\mathbf{R}_2)^{-1}(\mathbf{R}_1-\mathbf{R}_2)$ in order of decreasing eigen-values.

Efficient Implementation of QCFs

Consider now the problem of processing an entire 2D image using the quadratic filter. Let us assume that the test input vector **z** is obtained by lexicographically re-ordering the data in a relatively small window of row and column size $r\times c$ within a much larger image of size $R\times C$ [1]. The complexity of the direct computation necessary to evaluate the expression in Eq. 1 at all overlapping $r\times c$ sub-regions of the image is rather large. We therefore seek an efficient architecture for obtaining the full correlation of the input scene and the quadratic filter.

Denoting the dimensionality of the space by $d=r\cdot c$, let us assume that the $d\times d$ coefficient matrix $\mathbf{T}$ of rank N can be expressed as $\mathbf{T}=\mathbf{F}\mathbf{F}^T$, where $\mathbf{F}=[\mathbf{q}_1 \quad \mathbf{q}_2 \quad \cdots \quad \mathbf{q}_N]$ is a $d\times N$ matrix ($N\le d$). We note that the output of the quadratic filter can be expressed as

$$\varphi = \mathbf{z}^T\mathbf{T}\mathbf{z} = \mathbf{z}^T\mathbf{F}\mathbf{F}^T\mathbf{z} = \mathbf{v}^T\mathbf{v} \qquad \textbf{Eq 8}$$

where **z** is an r x c sub-region of the image and $\mathbf{v}=[v_1 \quad v_2 \quad \cdots \quad v_N]^T$ is a vector of the projection of **z** on the N columns of **F**. Clearly, **v** is a function of the spatial region of the image that is represented by **z**. The question is how the elements of **v** can be computed over the entire scene using efficient signal processing methods. Let $z(m,n)$ represent the full image to be processed. Furthermore, let us re-order the elements of $\mathbf{q}_i$ into a $r\times c$ mask $f_i(m,n)$. It is then easy to see that the values of v_i (i.e. i-th element of **v**) for all locations within the image can be obtained via the 2D correlation of $f_i(m,n)$ and input test image $z(m,n)$ as

$$v_i(m,n)=z(m,n)\otimes f_i(m,n), \qquad 1\le i\le N, \qquad \textbf{Eq 9}$$

where the symbol $\otimes$ indicates the 2D correlation operation. Here, we have expressed v_i along with the indices (m,n) to indicate that it is a function of the location of **z** in the image. The samples of $v_i(m,n)$ can be viewed as a partial result at each point in the image obtained by projecting the $r\times c$ region of the image centered at that point on $\mathbf{q}_i$. To obtain the quadratic term in Eq 8 for all points in the image, it now remains to square

[1] Typical sizes of real-world images may be 256 256 pixels (or larger) whereas the window size for **x** may be 20 x 20, chosen to be approximately the size of targets of interest.

the pixel surfaces of partial results and add them. Therefore, the output of the quadratic filter for all points in the image is given by

$$\varphi(m,n)=\sum_{i=1}^{N}|v_i(m,n)|^2=\sum_{i=1}^{N}|z(m,n)\otimes f_i(m,n)|^2 \qquad \textbf{Eq 10}$$

The expression in Eq 10 is a succinct way to express the output of the quadratic filter in response to the full input image. The computations can be readily dealt with using N 2D cross-correlation operations, each of which is efficiently implemented using fast Fourier transforms (FFTs). In fact, a more general form that takes eigenvectors with negative eigen-values into account is

$$\varphi(m,n)=\sum_{i=1}^{N_2}|z(m,n)\otimes f_i(m,n)|^2-\sum_{i=1}^{N_1}|z(m,n)\otimes g_i(m,n)|^2 \qquad \textbf{Eq 11}$$

where $f_i(m,n)$ and $g_i(m,n)$ are eigen-vectors corresponding to positive and negative eigen-values, each reshaped into an $r\times c$ mask.

Figure 1 depicts the architecture required to efficiently process a full image with the quadratic filter. Essentially, the input image is correlated with a bank of linear filters to obtain partial results that are squared and added to obtain the desired quadratic output for every point in the input image. Although several branches of linear filters are required, the architecture shown comprises a single quadratic filter implementation. Since each of the branches can be implemented using FFTs, the number of multiplications required in the architecture is proportional to N log(rc) where N is the number of branches, and $r\times c$ is the size of the filter masks. This QCF architecture is general in that any set of basis functions can be implemented, in this case specifically the TBF, not just those derived using the methodology outlined in this paper.

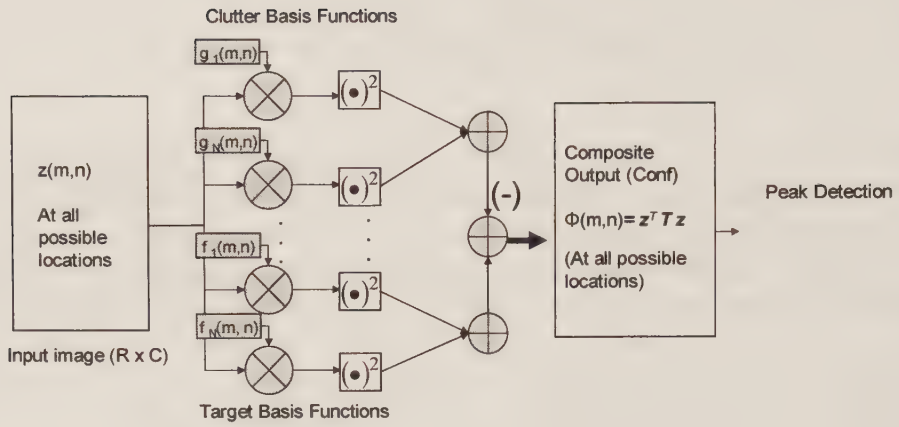


Figure 1 - QCF filter architecture

Detection metrics

Two detection metrics are used, the first being the peak-to-side lobe ratio (PSR) of the detected objects used for the MACH filter results. The PSR makes use of the standard z-score type normalization on the correlation surface where each PSR value is based on the local "target sized area" statistics as shown in Eq. 12, where μ is the local mean, σ is the local standard deviation, and τ is the original correlation value at each coordinate.

$$PSR \equiv (\tau - \mu)/\sigma \qquad \textbf{Eq 12}$$

The PSR is a measure of the sharpness of the correlation peak that has been shown to be maximized by the MACH filter design process [12], and is therefore a good metric to assess the performance of the MACH filter. On the other hand, PSR is not optimized by the QCF design algorithm, and is therefore not suitable for measuring QCF performance. For now, we use the composite output (a confidence value (Conf) created by the composite output of the architecture in Figure 1) as the QCF detection metric. In general however, it may be desirable to apply a normalization to the QCF output to allow meaningful comparisons across widely varying conditions.

The PSR and Conf measures produce results on different scales but since both show the relative detection values on the desired target and other objects, the gradient robustness can be discerned directly for comparison purposes.

The Initial Data set

Several samples of the initial data set are shown in Fig.2 which also points out the multiple targets in the sequence. Each additional test data sequence is similar but using different targets in different locations having consequently different backgrounds.

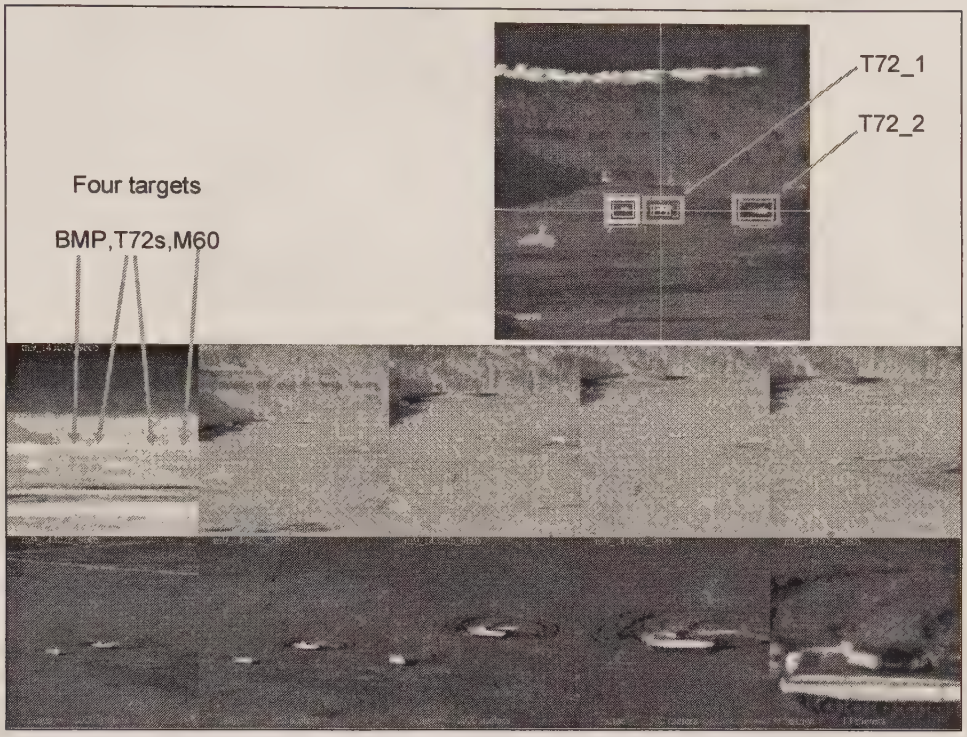


Figure 2 - Image sequence 1

Quantifying performance results

In order to illustrate the margin for discrimination we choose to use strip chart plots showing detection metric amplitude versus range-to-go which exemplifies the magnitude of the difference between the desired target delectability and other objects and non-desired targets which are potentially detected. We selected this type of chart since it depicts robustness in a much more intuitive way than typical receiver operating curves (ROC), which have been noted for misinterpretation [13].

Figure 3 shows the confidence value plotted on the TBF filter encoding of the primary T72_1 target (the T72 tank on the left in Figure 2), showing in red confidence values of the target detections and in black the other objects and clutter detected. At all ranges the target has a higher confidence value, however, between 4400 and 3600 meters there is reduced detection margin than ranges below 3500 meters until the filter is out of its useful range. The clutter is suppressed quite dramatically below 3500 meters range to go.

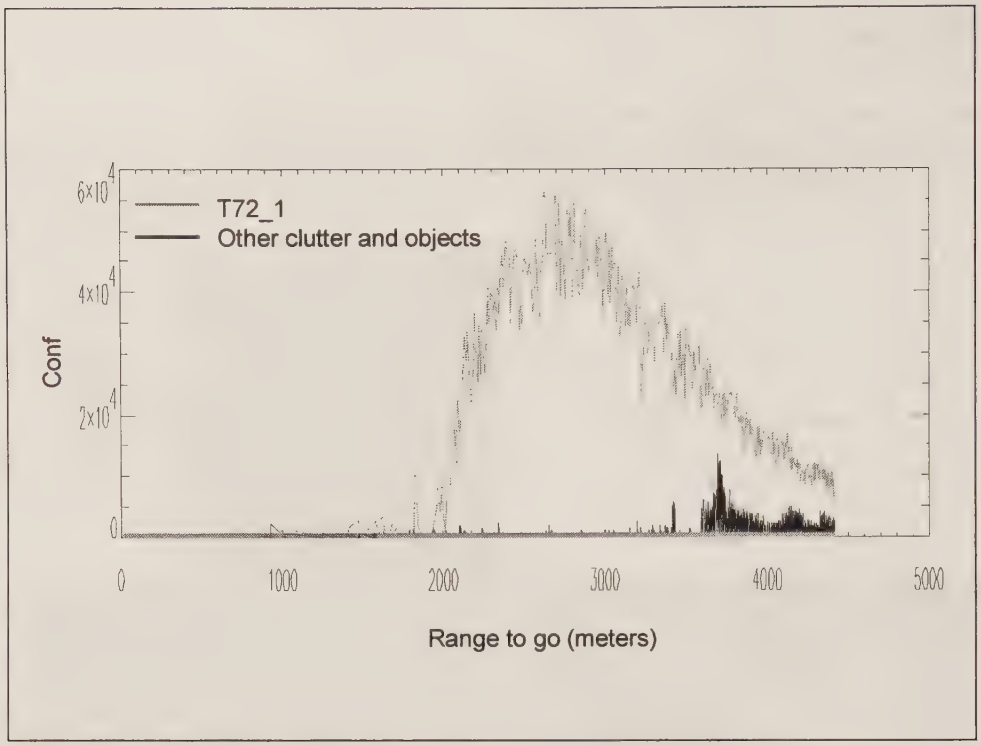


Figure 3 - TBF confidence of T72_1 trained from 4415 to 2695 meters

Figure 4, in contrast, shows the PSR value of the linear MACH filter results on three different MACH filters encoded with the same primary T72 target. Although the PSR and confidence value scales are different, it is clear the clutter detections are more prevalent and the discrimination margin is generally less and at some ranges non existent.

Figure 5 shows the confidence value plotted on the TBF filter encoding of the secondary T72_2 target (the T72 tank on the right in Figure 2), again showing in red the target detections and in black the other objects detected. The clutter at long ranges clearly has significantly higher confidence values. This T72 target goes out of the field of view at 2700 meters range to go, but note there is little clutter of consequence at the closer ranges.

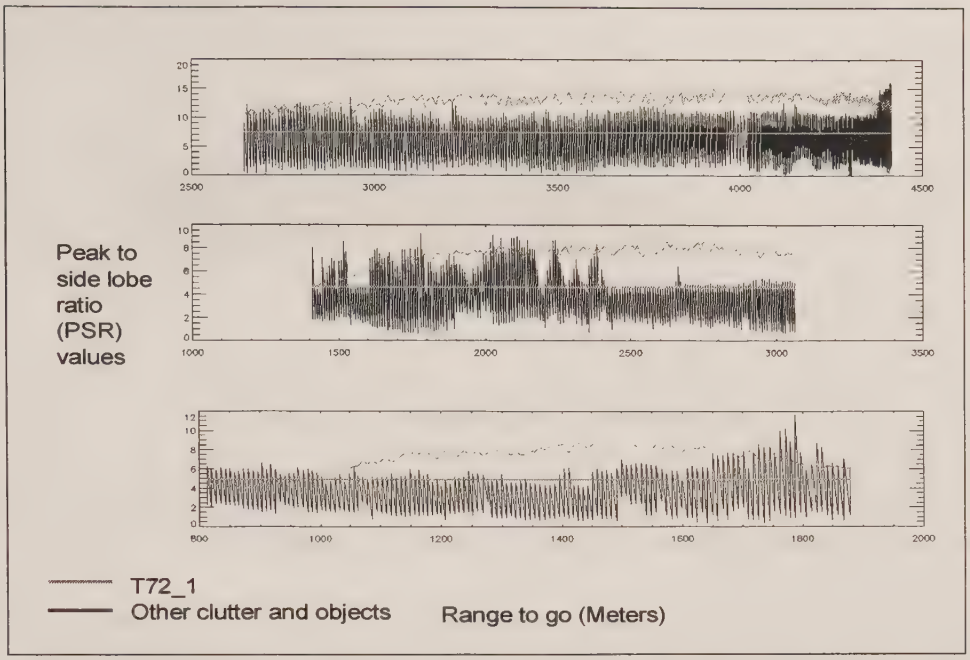


Figure 4- Three MACH filter results trained on T72_1 at their respective ranges

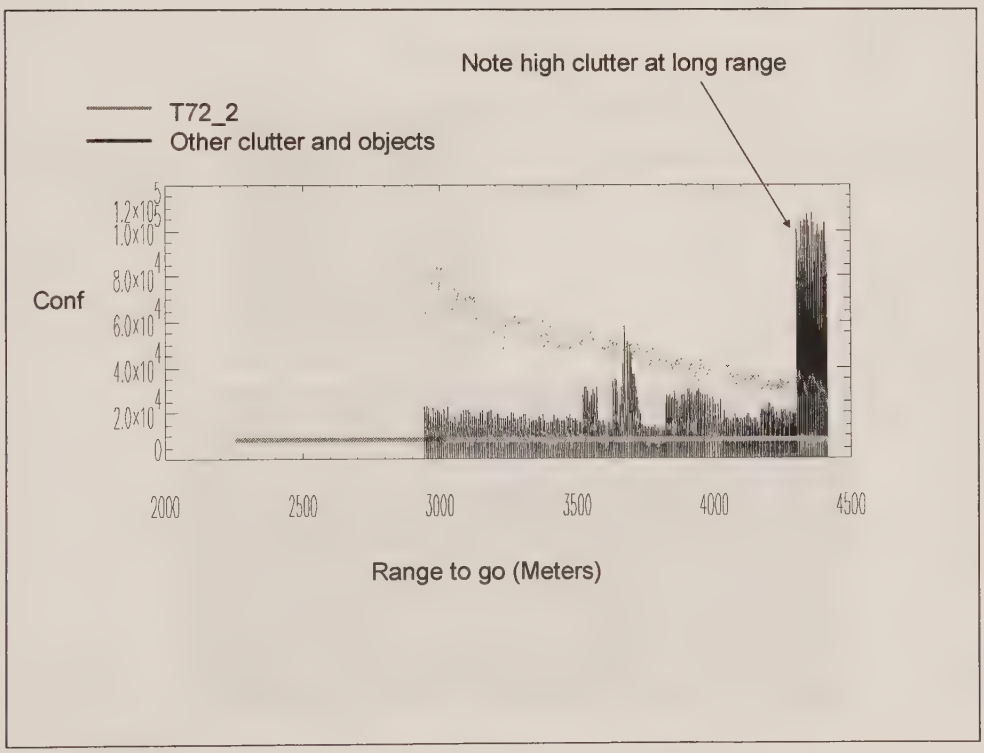


Figure 5 - TBF results on T72_2 target

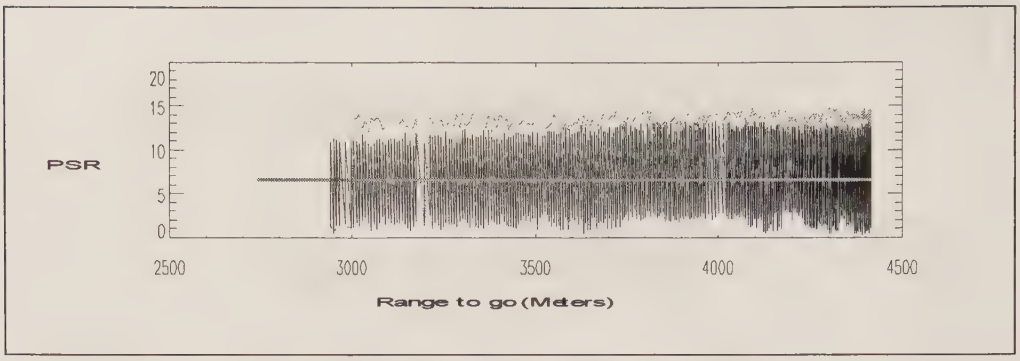


Figure 6 - MACH filter results on T72_2

Figure 6 shows the MACH PSR for comparison to the TBF results in Fig. 5. Again, there is little or no robustness margin in the MACH detection results.

Another data set with a different target is shown in Fig. 7. Note the TBF results in Fig. 8 have clutter confidence values that exceed the target value on several specific frames. These frames are unique in that the TBF results are distinctly different than the other frames due to a response to some noise. This noise can potentially be preprocessed out, however, it was left in here to emphasize the sensitivity of the TBF to certain noise.

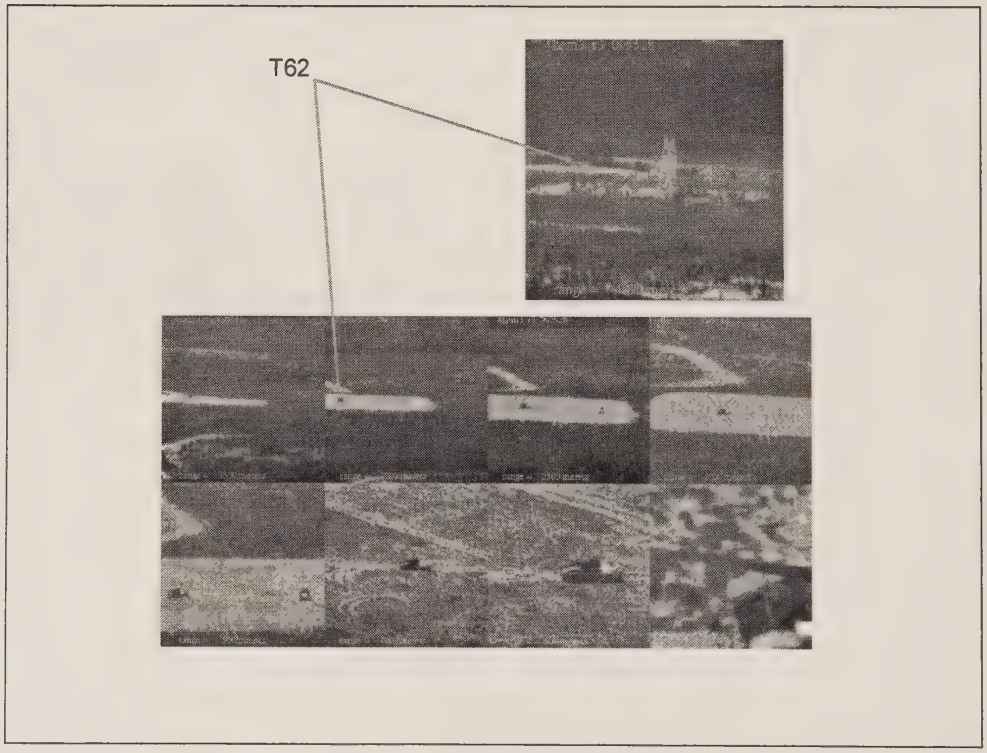


Figure 7 - Image sequence 2

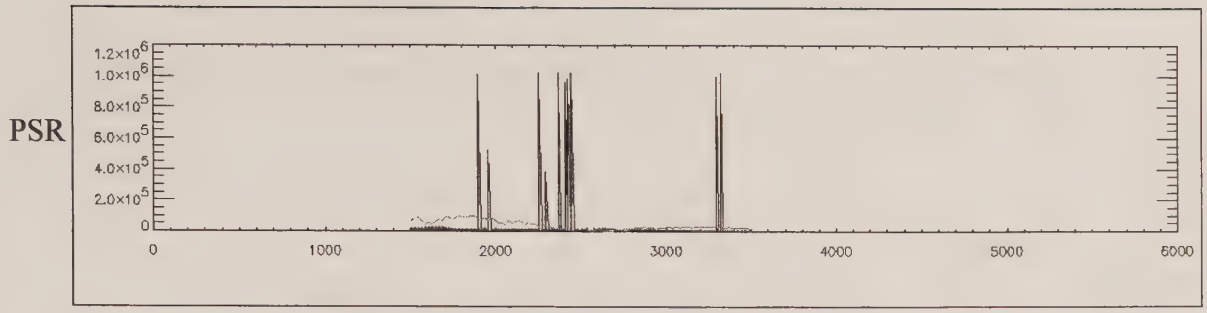


Figure 8 -TBF results on T62 (Image sequence 2)

Figure 9 is the same data that is plotted in Fig. 8 with the vertical scale changed such that the TBF target confidence value response is shown more clearly. Comparing Fig. 9 to Fig. 10 over the same range there is clutter/noise response which corresponds for both the TBF and MACH.

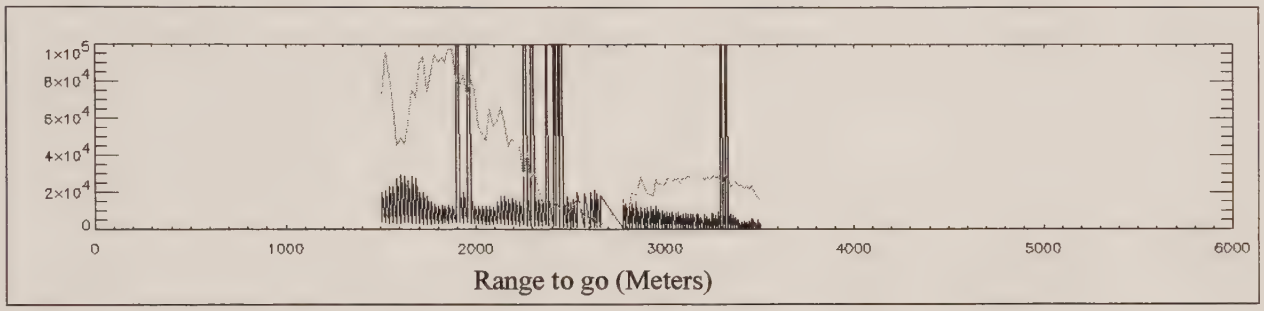


Figure 9 - TBF results on T62 (Image sequence 2)

The margin for discrimination with the TBF is shown to be better than the MACH overall on the T62 target, but, the MACH results show a healthy margin also.

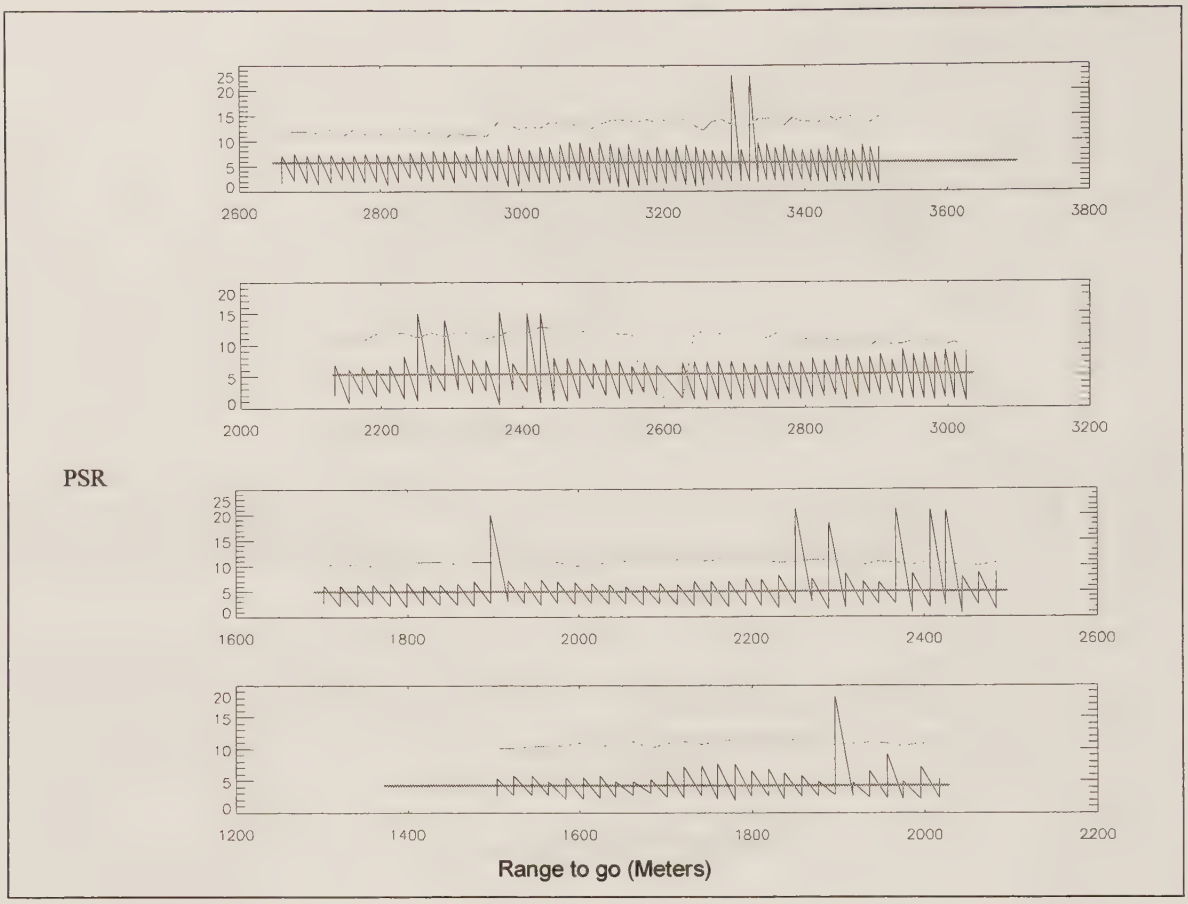


Figure 10 - MACH results on T62

Summary and Conclusions

In this paper, we have discussed the theory and implementation of QCF filters, and have shown performance on the Tuned Basis Function with strip chart plots delineating the robustness margin over varying range with several different data sets. We also quantified the MACH filter performance over the same data illustrating the difference in robustness with the TBF.

There are two main advantages of the QCF approach. First, there is only one composite correlation surface to be searched. This is simpler (both conceptually and computationally) than individually assessing multiple correlation surfaces created by a bank of linear correlation filters. The process of selecting the "best match" among the linear correlation surfaces can also be prone to errors. The second advantage is that a single QCF design can encompass much larger amounts of variations in the target signatures. Although the QCF's computational requirements are comparable to a bank of linear filters, the basis functions form natural partitions that are best suited for optimizing the QCF performance metric (Conf). With the TBF approach, a filter breakdown by range or orientation is not required or is certainly less constraining.

The comparison tests were staged such that the linear filters had every advantage, including smaller range requirements than the TBF filters. This was done to illustrate the robustness margin of the newer TBF filter encoding method. The TBF filters in every case have performed better consistently over the test data used, however, we have also shown cases where the TBF does not exceed the original linear MACH filters.

Future Work

We believe the test data provided some variation in target and background signatures, however, we plan continued tests on more varied data to fully characterize the detection performance for the TBF filters. Larger test sets that represent a broader range of conditions are essential for greater statistical significance, and a better understanding of the merits of the QCF detection approach.

The image pixel detection coordinates of the MACH and TBF have been shown to be different during our testing. Subsequent processing of the detected regions of interest using a classification post processor will help clarify any additional performance impacts of the TBF filters.

References

1. J. Starch, R. Sharma, S. Shaw, "A unified approach to feature extraction for model based ATR," *In the Proceedings of SPIE, Algorithms for SAR imagery III*, Vol. 2757, pp 294-305

2. D. Casasent, R. Shenoy, "Feature space trajectory for distorted object classification and pose estimation in SAR," Optical Engineering, Vol. 36, pp. 2719-2728, 1997

3. B. Bhanu, J Ahn, "A system for model-based recognition of articulated objects," *In the Proceedings of the International Conference on Pattern Recognition*, pp. 1812-1815, 1998

4. S. Z. Der, Q. Zheng, R. Chellappa, B. Redman, Hesham Mahmoud, "View based recognition of military vehicles in LADAR imagery using CAD model matching," Image Recognition and Classification, Algorithms, Systems and Applications, Editor B. Javidi, Marcel Dekker, Inc, New York, pp. 151-187, 2002

5. L.A. Chan, S.Z. Der, N.M. Nasrabadi, "Neural based target detectors for multi-band infrared imagery," Image Recognition and Classification, Algorithms, Systems and Applications, Editor B. Javidi, Marcel Dekker, Inc, New York, pp. 1-36, 2002.

6. D. Torreiri, "A linear transform that simplifies and improves neural network

classifiers," *In the Proceedings of the International Conference on Neural Networks*, 3, pp. 1738-1743, 1996.

7. H. S. Ranganath, D.E. Kerstetter, S. R. F. Sims, "Self Partitioning Neural Networks for Target Recognition," Neural Networks, Vol 8, No. 9, pp. 1475-1486, 1995 and Applications of Neural Networks in Electromagnetics, Editors C. Christodoulou and M. Georgiopoulos, Artech House, Inc. Boston – London, pp. 104 – 108, 2001.

8. B.V.K. Vijaya Kumar, "Tutorial survey of composite filter designs for optical correlators," Applied Optics, Vol. 31, pp. 4773-4801, 1992

9. A. Mahalanobis, B. V. K. Vijaya Kumar and S. R. F. Sims, "Distance classifier correlation filters for multi-class target recognition," Applied Optics, Vol. 35, 3127-3133 (1996).

10. A. VanderLugt, "Signal detection by complex spatial filtering," IEEE Transactions on Information Theory, Vol. 10, pp. 139-145 (1964)

11. A. Mahalanobis, B.V.K. Vijaya Kumar, S.R.F. Sims, J. Epperson, "Unconstrained Correlation Filters, " Applied Optics , Vol. 33, pp. 3751-3759, 1994.

12. A. Mahalanobis, B.V.K. Vijaya Kumar, "On the optimality of the MACH filter for detection of targets in noise" *Optical Engineering*, Special Issue on Correlation Pattern Recognition, Vol. 36 (10), pp. 2642-2648, October 1997

13. S. R. F. Sims, "Putting ATR performance on an equal basis: The measurement of knowledge base distortion and relevant clutter," *In the Proceedings of SPIE, Automatic Target Recognition X*, pp.55-60, Vol. 4050, April 2000.

Performance of multi level error correction in binary holographic memory

Jay C. Hanan*, Tien-Hsin Chao, George F. Reyes

Bio-Inspired Technologies and Systems, Jet Propulsion Laboratory, CA, USA

ABSTRACT

At the Optical Computing Lab in the Jet Propulsion Laboratory (JPL) a binary holographic data storage system was designed and tested with methods of recording and retrieving the binary information. Levels of error correction were introduced to the system including pixel averaging, thresholding, and parity checks. Errors were artificially introduced into the binary holographic data storage system and were monitored as a function of the defect area fraction, which showed a strong influence on data integrity. Average area fractions exceeding one quarter of the bit area caused unrecoverable errors. Efficient use of the available data density was discussed.

Keywords: error correction, holographic memory, data storage, parity, bit matrix, particle analysis

1. INTRODUCTION

1.1 Holographic Memory

A driving force improving digital storage technology is data density. Given equal data transfer rates and environmental compatibility, the more bits/volume the better the device. The theoretical data density limits of holographic memory, where the data may be stored in a volume, far exceed the theoretical data density limits of current technology such as magnetic disks, where information is confined to a surface area.

The Jet Propulsion Laboratory (JPL) is currently developing a new holographic memory system with performance characteristics including: read/rewrite capability, high-density, high transfer rate, non-volatility, compactness, and radiation resistance. These characteristics are selected to meet requirements for data storage needs for both NASA's space missions and commercial applications.[1,2,3,4,5,6,7,8,9,10]

NASA's future missions would require massive high-speed onboard data storage capability to support Earth Science missions. With regard to Earth science observation, a 1999 joint JPL and Goddard Space Flight Center (GFSC) study[3,4] ("The High Data Rate Instrument Study") has pointed out that the onboard science data (collected by high date rate instruments such as hyperspectral and synthetic aperture radar) stored between downlinks would be up to 40 terabits (Tb) by 2003. However, onboard storage capability in 2003 is estimated at only 4 Tb which is only 10% of the requirement. By 2006, the storage capability would fall further behind and would only be able to support 1% of the onboard storage requirements.

The volume nature of holographic data lends itself to a form of sectoring data structures across pages. Each page is given an angular address and contains a two-dimensional array of data. At JPL a binary holographic data storage system (BHDS) was constructed and tested with methods of recording and retrieving the binary information. Levels of error correction were introduced to the system and their performance is reported here.

1.2 Error Correction

An example of standard error correction in memory applications requires the dedication of a bit to each byte that may be checked. This "9th" bit may be assigned a value based on the sum of the byte, such as 'on' for odd and 'off' for even—parity. Checking parity allows errors in data to be discovered and, in some cases, corrected. For Holographic Memory, a similar error correction method may be employed. Every 9th bit may be dedicated to checking the sum of the previous 8

*Corresponding author; electronic mail: Jay.C.Hanan@jpl.nasa.gov; phone 818-354-4820

Optical Pattern Recognition XV, edited by David P. Casasent, Tien-Hsin Chao, Proc. of SPIE Vol. 5437 (SPIE, Bellingham, WA, 2004) · 0277-786X/04/$15 · doi: 10.1117/12.548068

bits, however since the data is stored in pages and may be represented as a binary matrix, other more efficient error checking methods may be employed.

For the experimental system described below, the most basic form of error correction (or prevention) is to increase the signal per bit—a form of bit averaging. As this is a costly form of error correction, its use can be minimized to keep data density large. However, in cases were transfer speed and data integrity outweigh information density concerns, the ability to dynamically alter the bit averaging level can play a valuable role.

A more efficient error checking scheme increases the data density without substantial loss of data integrity. Selecting this scheme requires consideration for the hardware peculiarities and the data format.[11] For example, based on measured dynamic range, the data stability is typically greatest toward the geometric center of the page where beam flatness is greatest; thus a measurement of errors toward the center of a page will provide less information on the likelihood of errors toward the edge of a page. If we assume the converse is true, then a reduction in the number of checksums and a corresponding increase in data density may be realized through an edge weighted page based error checking method. A simple page based error checking method is outlined in Figure 1 below.

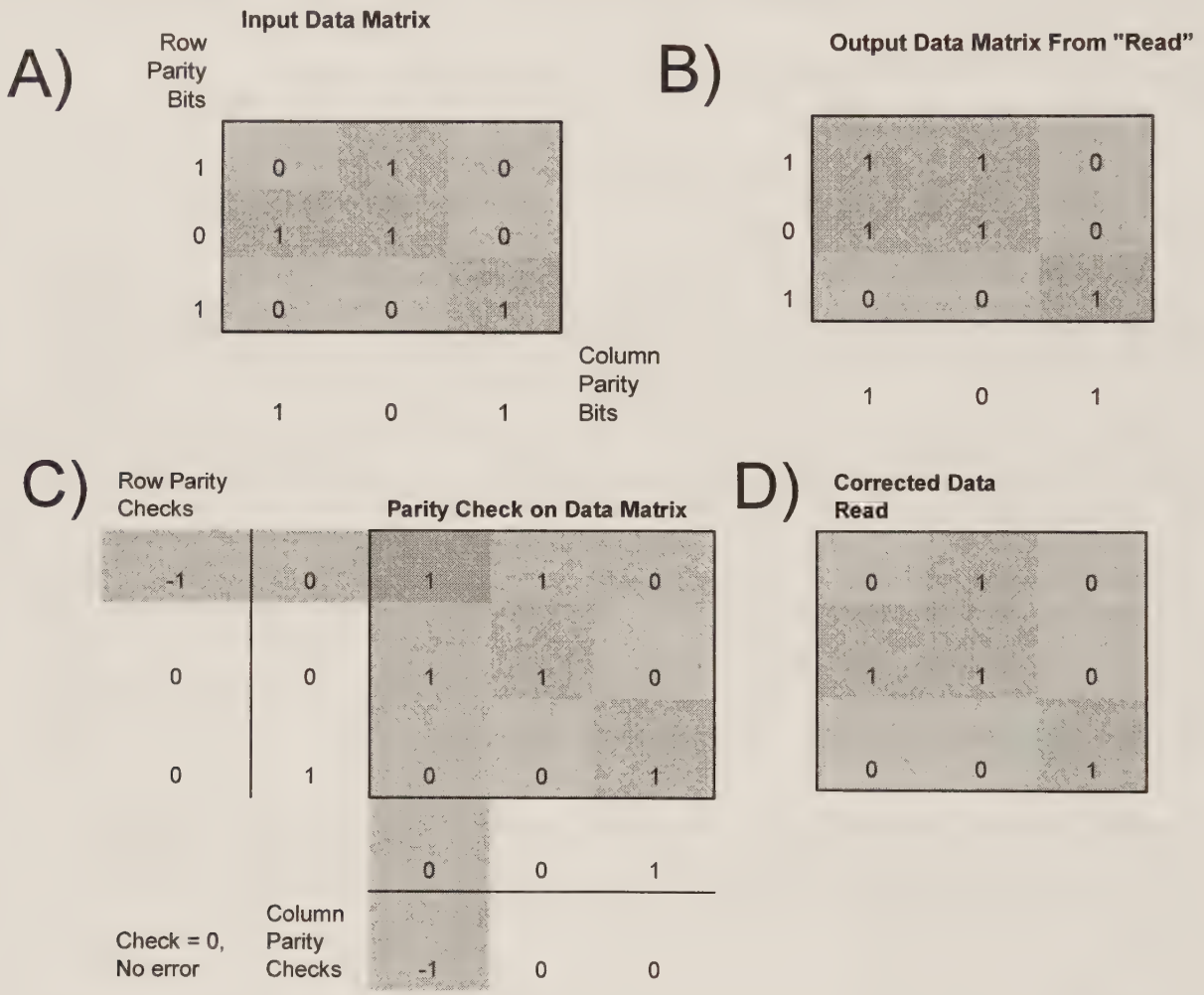


Figure 1. A schematic diagram of error correction using the L-Code on a 3x3 matrix. Extension to larger data matrices provides a simple error correction code first tested with the holographic system.

In the example above, an input data matrix of 3 x 3 bits along with its parity bits is displayed at A). In B) an example read is demonstrated with an error in the upper left bit. The error is discovered in C) with the parity check. In D) the error is corrected without re-reading the data. This scheme extends to N x M page matrices. The N x M matrix can be

subdivided for optimal data density under expected noise conditions. Further, introducing a whole page parity check provides an additional data verification step. Since the parity bits lie along one column and row, they can be placed in the form of an "L"; hence this method of error correction will be referred to as the L-code.

2. METHODOLOGY

2.1 Laboratory System

The current laboratory setup allows analysis of all bits to determine the expected noise characteristics of the BHDS. The analysis is performed using custom LabVIEW based analysis software. The arrangement of parity bits will be determined based on analysis of these noise characteristics. Beam steering technique, image sensor performance, and other components factor into the total noise characterizes of the final system. Further evaluation of improved components will continue as the technology develops.

The BHDS system has two main branches. One branch is used for retrieving stored information (read) and both branches are used for storing holographic information (write). The distinctive components of the read branch are the multiplexing device and the photo detector for a data input component. While several technologies may be used for multiplexers, the BHDS uses a piezoelectric steerable mirror. For detailed error analysis a charge coupled device (CCD) was used to retrieve holograms from the photorefractive crystal. While this analysis employed an Fe doped $LiNbO_3$ crystal, the results are applicable to other storage media. Along with intersecting the photorefractive crystal writing holograms, the second branch includes a spatial light modulator (SLM) in order to impart the page information onto the source laser beam (see the labeled photograph Figure 2).

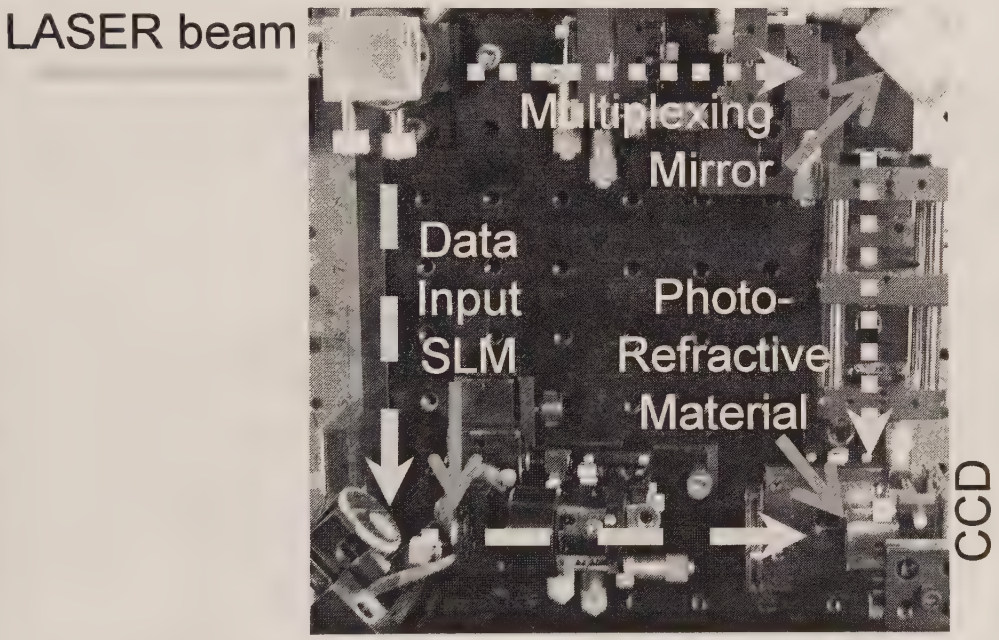


Figure 2. Photograph including labeled components of a Binary Holographic Data Storage (BHDS) system. The laser beam marked by green arrows is split into two branches. For reading data, only one branch (short dash) is on. For writing, both branches are on creating a hologram of the SLM plane with two beam interference in the photorefractive material.

2.2 Experimental Procedure

A method was developed for testing error tolerance in the system. In the case of field use, the source of errors may be unknown. An advantage of holographic memory is the capability to dynamically reconfigure the page to compensate for defects. Thus, in the laboratory, sources of error are typically minimized or eliminated. However, for this analysis, errors were intentionally introduced during the write procedure to test the multiple forms of error correction.

Errors were simulated in the device using a damaged SLM. Damage to the SLM included 45 spots, some minor scratches and other defects disrupting the intensity from the SLM. The same defect plane was used for all measurements. Since the size and arrangement of the defects were constant, the effect of defect area fraction was investigated by adjusting the area/bit. The area of bits was increased by incrementing the number of SLM pixels/bit, n x n pixels/bit where, $n = 2, 3, 4, 5, 6$, and 7. Thus, five separate bit areas were tested against the constant defect plane giving five average defect area fractions, [average defect area] / [bit area] (See Table I). This increase in bit area was expected to reduce the number of observed errors—a first level of error correction in the device.

A second, more efficient level of error correction is available through the use of parity checks (see §1.2). Two error correction methods were used on the page arrays. The first was the simple L-code error correction method using only one "L" per page. The second is a more efficient method developed at Caltech called the X-Code.[12] Both of these codes were tested on a 640x480 pixel^2 representative portion of the data plane at each value of n. This format coincides with the format of the CCD mounted to the BHDS for purposes of testing. While the full data density of the system cannot be read by the 640x480 pixels^2 CCD, the number of bits observed at each scale provides representative area of the page for analysis.

For the purpose of this experiment, bits were defined as on if their average intensity/bit exceeded the threshold value, 800 out of 1024 (10 bit CCD), where the source laser power was optimized for best use of the available dynamic range. In practice, an optimal threshold value depends on several factors affecting the background intensity including the number of pages recorded. In practice a dynamic threshold method can be employed to further improve error tolerance; constant thresholding was used in this case to enable a germane comparison between the observed bit areas.

3. RESULTS

As expected more errors were observed as n decreased (see Table I). For $n = 7$ no error bits were observed. The number of error bits remained small down to $n = 4$. At $n = 3$ a significant increase in error bits was observed concluding with the greatest number for $n = 2$. Both the L-Code and the X-Code were able to recover data from errors up to $n = 4$. At $n = 3$, where on average 31% area of the affected bits are obscured by defects, the L-Code was able to recover more than half of the errors and the X-Code nearly three-quarters (Table I).

Table I. Results from measuring error bits for each pixel area (n x n).

n (pixels)	Defect Area Fraction	Raw Reconstruction (Error Bits/Total Bits)	L-Code Correction (Improvement)	X-Code Correction (Improvement)
7	8%	None Detected	-	-
6	10%	0.22%	100%	100%
5	14%	0.11%	100%	100%
4	20%	0.13%	100%	100%
3	31%	0.87%	52%	74%
2	56%	1.13%	6%	48%

As n decreases and the error density increases and clustering of errors became more frequent. These clustered errors cannot be corrected by the L-Code. However, the number of errors increases non-linearly with defect area fraction to the degree that both error correcting codes failed at the same $n \leq 3$ (see **Figure 3**).

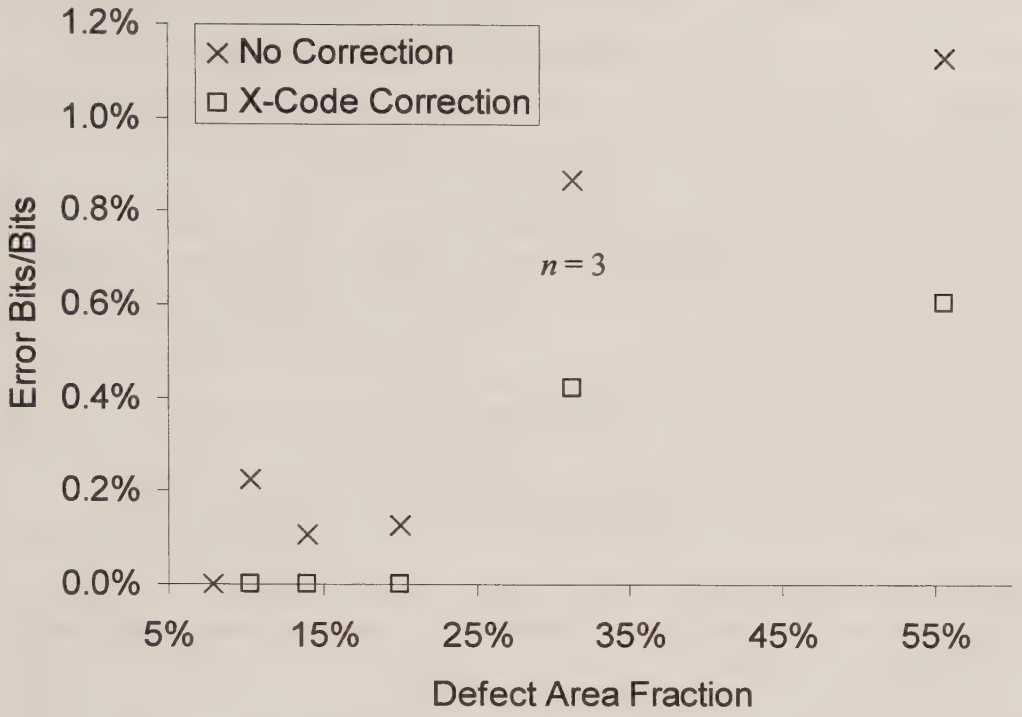


Figure 3. Graph depicting the increase in error bits as a function of increasing defect area fraction. Error correction from both the X-code (shown) and the L-Code (not shown) reduces the error to a non-detectible limit for defect area fractions of 20% or less.

4. CONCLUSIONS

Errors were artificially introduced into a binary holographic data storage system (BHDS) constructed at the Jet Propulsion Laboratory. Data was stored in binary page format to the BHDS. The intentionally introduced read errors were monitored as a function of the defect area fraction. Defect area fraction showed a strong influence on data integrity with average area fractions exceeding one quarter of the bit causing unrecoverable errors—regardless of tested error correction schemes. Operation close to this one quarter level provides the most efficient use of the available data density. Successful data storage and recovery was performed at this level. Of the two error codes tested the X-Code corrected more errors suggesting the X-Code provides a greater margin of safety for data integrity; however, both codes tested provided similar practical performance.

ACKNOWLEDGEMENTS

The research described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

REFERENCES

1. T. H. Chao, G. F. Reyes, J. C. Hanan, H. Zhou, D. Dragoi, "High-Density Holographic Memory" *ISOM* (2001).
2. T. H. Chao, J. C. Hanan, G. F. Reyes, "High density high data rate holographic memory using a MEMS mirror beam steering device." *NASA NTR-40785* (2003).
3. K. F. Strauss, T. Daud, "Overview of radiation tolerant unlimited write cycle nonvolatile memory," *Proc. 2000 IEEE Aerospace Conference*, vol. 5, pp. 399-408, 18-25 March (2000).
4. W. Schober, F. Lansing, K. Wilson, E. Webb, "High Data Rate Instrument Study", *JPL publication 99-4*.
5. T. H. Chao, H. Zhou, G. F. Reyes, JPL "Compact Holographic Data Storage System,", *Proceedings of Eighteenth IEEE Symposium on Mass Storage Systems in cooperation with the Ninth NASA Goddard Conference on Mass Storage Systems and Technologies*, April. (2001).
6. T. H. Chao, H Zhou, G. F. Reyes, "Advanced compact holographic data storage system," *Proceedings of Non-volatile memory technology symposium 2000*, pp. 100-105, November, (2000).
7. T. H. Chao, G. F. Reyes, H. Zhou, D. Dragoi, J. C. Hanan, "High-density Holographic Data Storage," *Proceedings of International Symposium on Optical memory 2001* PP.248-249, Taiwan, Oct. (2001).
8. T. H. Chao, G. F. Reyes, H. Zhou, D. Dragoi, J. C. Hanan, "Nonvolatile Rad-Hard Holographic Memory," *Proceedings of Non-volatile memory Technology Symposium 2001*, pp. 12-17, San Diego Ca, Nov. (2001).
9. T. H. Chao, "Advanced Holographic Memory_AIST-99-0057_Final Report" *Prepared for NASA ESTO AIST*, Nov. (2002).
10. D. Dragoi, T. H. Chao, "Radiation test of LiNbO3 photorefractive crystal," *JPL report* (2002).
11. H. Pishro-Nik, N. Rahnavard, F. Fekri, "Non-uniform Error Correction Using Low-Density Parity-Check Codes" *IEEE Transaction on InformationTheory* (2003).
12. L. Xu, J. Bruck, "X-Code: MDS Array Codes with Optimal Encoding." *IEEE Trans. on Information Theory*, Jan. (1999).

Photoconductive Optically Driven Deformable Membrane for Spatial Light Modulator Applications utilizing GaAs and InP Substrates

B. Haji-Saeed, R. Kolluru, D. Pyburn, R. Leon, S. K. Sengupta, M. Testorf, W. Goodhue
Physics and Electrical and Computer Engineering Departments,
University of Massachusetts Lowell, Lowell MA 01854

J. Khoury, A. Drehman and C. L. Woods
Air Force Research Laboratory / SNHC, Hanscom Air Force Base, MA 01731

J. Kierstead
Solid State Scientific Corporation, Hollis NH 03049

ABSTRACT

The fabrication and characterization of an optically addressable deformable mirror for spatial light modulator is described. Device operation utilizes an electrostatically driven pixellated aluminized polymeric membrane mirror supported above an optically controlled photoconductive GaAs substrate. A 5-μm thick grid of patterned photoresist supports the 2-μm thick aluminized Mylar™ membrane. A conductive ZnO layer is placed on the backside of the GaAs wafer. Similar devices were also fabricated with InP. A standard Michelson interferometer is used to measure mirror deformation data as a function of illumination, applied voltage and frequency. A simplified analysis of device operation is also presented.

Keywords: MEMS, Deformable Mirrors, Spatial Light Modulator.

1.INTRODUCTION

Deformable mirror spatial light modulators are the key enabling components of many optical signal processing systems including optical correlators and adaptive optics systems for atmospheric turbulence correction. Such mirrors may be formed of many independently actuated rigid sections (segmented mirrors) or with large area reflective flexible membranes (membrane mirrors) with suitable actuators. Currently, there are three main technologies for addressing the membrane mirror light modulation (MLM) [1]. This includes electron beam addressing [2,3], optical addressing [4-7] and electrical addressing via integrated circuits [8-10]. Membrane mirror spatial light modulators have been used as large-screen display systems [2], infrared scene projectors [3] and light shutters [1]. In the first two designs [2,3] there is a charge-transfer-plate (CTP), which is addressed by the electron-beam in the back, and the reflective membrane is suitably supported on the front side. Wavefront correction using MEMS Micro-mirrors have been done recently, utilizing a sophisticated and complicated four-layer planarized SUMMiT4 process for fabrication [11].

We propose and demonstrate a new deformable mirror device for use in adaptive optics and optical signal processing. Our approach is based on an optically addressable deformable mirror spatial light modulator. Device processing is simple, in that only one mask level is required, at least for prototype devices. Instead of difficult to fabricate CTPs with conductor pins inside an insulating matrix, we use a single semi-insulating GaAs or InP wafer. An infrared (IR) transparent conducting ZnO layer is deposited on the backside of the wafer using a special process developed at AFRL. These materials allow us to work at the shorter infrared atmospheric windows. Unlike ITO, which is reflective in the infrared, our ZnO coating is transparent to several microns in wavelength. GaAs and InP are photoconductive, creating a device that can be addressed optically. The III-V materials when suitably doped, exhibit photorefractive behavior, which allows the possibility of holographic addressing which in turn opens up new opportunities in adaptive image correction. Instead of using an electron beam for addressing, we utilize laser beam illumination, which creates a simpler device that does not require a high vacuum for operation. Using aluminized Mylar® as the membrane is another notable feature of our design. Mylar® is an inexpensive, thin and very elastic material. Precise deformation measurements of fabricated devices are important for their characterization. Quan et al. have utilized laser interferometry for the deformation

Optical Pattern Recognition XV, edited by David P. Casasent, Tien-Hsin Chao, Proc. of SPIE Vol. 5437 (SPIE, Bellingham, WA, 2004) · 0277-786X/04/$15 · doi: 10.1117/12.548069

measurement of a micro-rf switch membrane. This method is based on a modified Michelson interferometer. A DC voltage is applied to the device and the results show the membrane deformation and the applied voltage relationship [12]. We study our device similarly using a standard Michelson interferometer for deformation measurement as both AC and DC voltages are applied. The capture and computer processing of the entire interferometric fringe pattern allows rapid evaluation of the entire device area. In addition, single pixels could also be carefully monitored. We have also used more traditional measurement techniques utilizing a single photodetector for monitoring fringe shifts for the measurement of device responses.

2.DEVICE STRUCTURE

Figure 1 shows the architecture of our device. It is fabricated on a 2-inch semi-insulating GaAs wafer. An IR transparent conducting ZnO layer is deposited on the backside. We have also used InP wafers. A 5-micron thick grid of photolithographically patterned developed photoresist on the front-side serves as the membrane support. A 2-micron thick aluminized Mylar™ membrane is attached to the support to complete the device. The device operates as follows: a bias voltage is applied between the membrane and the back contact. When a light is shone on the back of the device, the conductivity of the GaAs layer increases. This leads to a higher effective electrostatic drive between the suspended membrane and the front-side of the semiconductor, causing increased membrane deflection in the illuminated areas.

3.DC BIAS THEORY

We provide here a summary of the simplified basic analysis of device operation. A more detailed treatment will be published elsewhere. During operation, a DC voltage is applied to the device. The equivalent circuit of each pixel of the device may be described as two capacitors C_1 and C_2 in series, where they are the capacitances associated with the photoconductive substrate and the membrane-substrate interface respectively. This model is illustrated in figure 2. After light illumination the device acts as a series RC circuit, shown in figure 3, in which the membrane remains as a capacitor but the photoconductor changes to a resistor. In this case, in the steady state, the applied voltage drop is just on the capacitor (membrane).

From simple electrostatic considerations, based on earlier work [2,3], the change in deflection upon illumination can be written as:

$$\Delta h = \frac{\varepsilon_0 r_1^{\,2} V^2}{8Ts^2}\left(\frac{C_1^{\,2}}{(C_1 + C_0)^2} - 1\right) \qquad (1)$$

where C_0 is the capacitance of the membrane when there is no deflection (h=0). From this equation, under DC bias operating conditions $C_0 < C_1$, therefore there is only a small deflection of the membrane. Here, V is the applied voltage, ε_0 is permitivity of free space, $2r_1$ is the size of one square of the pixel pattern (the experimental devices utilize square pixels, except for the single pixel device, while we assume circular geometry pixels for the analysis itself), s is the depth of the well and h is the deflection of membrane, and T is the tension on the membrane. It is important to note that the deflection is independent of light intensity – only two states of deflection corresponding to the light being on or off are possible. That is, binary operation is achieved with optical addressing.

4.FABRICATION

The test devices were fabricated as follows. Two-inch diameter semi-insulating GaAs wafers (0.50 mm thick) were coated with a thin conducting layer of ZnO on the backside by a proprietary process. The ZnO layer thickness varied from 310 Å to 630 Å from wafer to wafer. A thick layer (~5-μm) of Clariant AZ-4620 positive photoresist was deposited on the front side by spin coating and patterned using photolithography to create the spacer layer. Chromium on glass photomasks were used for the 1mmx1mm and 2mmx2mm samples. An ordinary paper punch was used to create a 7 mm diameter hole in aluminum foil for the 'punch hole' mask. A 2 μm thick Mylar® membrane was coated with about 4000 Å of aluminum by E-beam evaporation.

The uncoated side of the Al coated film was attached to the top of the photoresist layer by a thermal bonding process at 110-120°C. Due to the heat, the photoresist becomes adhesive and the polymer film adheres to it.

5.CHARACTERIZATION

We used a Michelson interferometer for the characterization of the devices that we have fabricated. Three different samples have been characterized. Back illumination was provided by a 125-W lamp and a 50-mW diode laser (wavelength λ=980 nm).

Detailed measurements on single pixels could be made by focusing the interferometer beam onto the device, with a lens, so that only one pixel was illuminated. Quantitative measurements were obtained by monitoring fringe shifts with a photodetector. Fringe images were captured with a CCD camera coupled to frame grabber interfaced to a computer.

5.1. Sample (wafer) characterization

We measured the IR absorption spectra of the GaAs (Gallium Arsenide) and InP (Indium Phosphide) wafers for selecting a convenient wavelength for optical addressing. The actual spectra will be published elsewhere. We expect, light in the moderately absorbing region near the band edge wavelength, to be the most effective in increasing the bulk electrical conductivity of the semiconductor, and thus to be optimal for optical addressing. Shorter wavelengths would be absorbed at the surface and longer wavelengths would generate few mobile charge carriers. The absorption spectra showed that for GaAs, the optimal wavelength is around 0.89 μm (890 nm). Unfortunately, a suitable laser source was not available at this wavelength. We use either a broadband lamp source, or a 980 nm source to demonstrate optical addressing.

5.2. Device optical characterization

A standard Michelson interferometer was used to monitor the physical deformation of the membrane mirror as a function of operating parameters such as bias voltage and back illumination. It also allowed us to monitor the optical quality of the membrane surface in both the unenergized and biased state, which is important in wavefront correction applications. Figure 4 shows our experimental arrangement. An expanded and collimated 5-mW HeNe laser beam (7mm in diameter) is split via beam splitter BS_1 (25mm x 25mm) into two beams corresponding to the two arms of the interferometer. One beam impinges upon a plane mirror mounted on a PZT linear translator driven by either DC or sawtooth voltages. This piezo-driven mirror helps us to adjust the interference in the linear operational range and also we can calibrate our system by adjusting the PZT until we get the maximum fringe shift. The other beam impinges upon the MEMS device. An AC or DC voltage is applied to the MEMS. We chose to limit the drive to the device to keep us within the linear range of the interference. We adjusted both the MEMS drive voltage and the DC voltage to the PZT to get a sinusoidal output signal.

Beam splitter BS_2 is used for splitting the beam into the photodetector and the CCD camera. In some experiments the beam is passed through an iris to select a single fringe before imaging by the camera. The beam to the detector is always passed through a second iris to select just one fringe and be on the linear range or on the central fringe for straight line or circular fringes respectively. The photodetector output is connected to an oscilloscope and the CCD camera sends the data to a frame grabber and frame grabber transfers the data to a computer.

Figure 5 show the results of our experiment for the punch hole sample, which is a large circular single pixel device with a pixel diameter of 7 mm. Both the single fringe from a focused beam on the sample and the entire fringe pattern of the punch hole are shown. Figure 5 demonstrates the fringes before [5(A, A')] and after [5(B, B')] the light illumination with the lamp on the back of the sample. The difference between these two sets of figures, produced by computer subtraction to show us the movement of the fringe, is also presented. The unprimed images 5(A, B, C) are the focused area results, while the primed images 5(A', B', C') are the broad area measurements.

We studied the effects of the pixel size on the response of the devices. Figure 6 shows the 1mm x1mm pixel sample, and the 2mmx2mm pixel sample, before [6(A, A')] and after [6(B, B')] laser and lamp illumination respectively, on the back of the sample, and the corresponding difference images. An examination of the fringe difference patterns in Fig. 6(C') indicates that the deflection of the active pixels is about 1/8 λ of the He-Ne wavelength, which in agreement of our theory, is small. There are no straight-line fringes in either of the two samples because the membranes are not perfectly flat either in the undeflected or deflected state. Figure 7 shows the fringes focused on one pixel of the 1mmx1mm device, before 7(A) and after 7(B) light illumination on the back of the samples and the difference image 7(C).

Figure 8(A) shows the resulting deformation of the membrane as a function of the amplitude of the applied AC voltage for the 1mmx1mm pixel sample at 4 Hz. These measurements are derived from single photodetector studies of fringe shifts. The deformations are expressed in terms of the peak-to-peak output voltage of the photodetector monitoring the fringe shifts. Each unit of the displacement scale corresponds to a surface deflection of ¼ wavelength of the He-Ne wavelength or about 0.16 μm. The figure shows the experimental data and the best-fitted curve. We expect that the curve for the deformation of the membrane as a function of voltage should be roughly parabolic as the electrostatic force for a fixed capacitance device is proportional to the square of the voltage. The more detailed analytical model also supports this idea. Figure 8(B) demonstrates the relationship between the membrane deformation and the frequency with 30V applied, for the 1mmx1mm sample. The membrane mirror system behaves both electrically and mechanically as a low pass filter. Figures 9(A) and 9 (B) similarly demonstrate the relationship between the membrane deformation and the drive voltage and the drive frequency for the 2mmx2mm sample.

The results of membrane deformation for the punch hole sample vs. both frequency and voltage are shown in figures 10(A) and 10(B). These measurements were made without any back illumination. Essentially identical response was observed under illumination. The frequency response is sufficient for many adaptive optics applications. Recent designs are of much higher sensitivity and will be reported elsewhere.

6.CONCLUSIONS

We have proposed and fabricated a new deformable mirror spatial light modulator (MLM), which is optically addressable. Optical addressing has been done by a 125 W lamp and a 50-mW diode laser (wavelength λ=980 nm).
The device is an aluminized membrane mirror supported by a grid of developed photoresist in proximity to a photoconductive GaAs or InP substrate with an IR transparent conducting ZnO backside electrode. A standard Michelson interferometer measured the deformation and performance of our device as a function of illumination, applied voltage and frequency. The basic concept of the optical addressing of a deformable mirror was verified. The devices show a binary response to optical illumination and acceptable frequency response. Further work is indicated.

ACKNOWLEDGMENTS

We would like to thank Lionel Bouthillette for the spectrophotometric measurements. Helpful discussions with Michael Alexander on photoconductivity in GaAs are hereby acknowledged.

REFERENCES

1. C.Warde, J.McCann, J.MV.Shrauger, H.Ieong, A.Ersen, X-YWang and J.Hubbard "Membrane Mirror Light Modulator Technology," Proceedings of SPIE 191,Vol.3951 (2000).
2. C.Warde, J. E.Hubbard, Jr., G. Genetti, L.Lerman, W.Loizides "Membrane Mirror Light Valve for High-Definition Projection Display," Proceedings of SPIE 270,Vol.1910 (1993).
3. C.Warde, J. E.Hubbard, Jr." Membrane -Mirror- Light -Valve-Based Infrared Scene Projector," Proceedings of SPIE 544,Vol.2223 (1994).
4. L.E. Somers, "The photoemitter membrane light modulator image transducer," in Advances in Electronics and Electron Physics, Vol.33A (Academic Press, New York, 1972), 493.
5. F. Reizman,"An Optical Spatial Phase Modulator Array Activated by Optical Signals," in Proceedings of the 1969 Electro-optical Systems Design Conference, pp225-230.
6. A.D.Fisher, L.C. Ling, J.N.Lee and R.C. Fukuda,"Photoemitter Membrane Light Modulator," Opt. Eng. 25, 261-268, 1986.
7. D.R. Pape,"Optically Addressed Membrane Spatial Light Modulator," Opt.Eng. 24, 107-110, 1985.
8. K.Preston, Jr.,"An Array Optical Spatial Phase Modulator," Proceedings of the IEEE International Solid State Circuits Conference, NewYork, 1968, p.100.
9. D.R. Pape and L.J. Hornbeck,"Characteristics of the Deformable Mirror Device for Optical Information Processing," Opt.Eng. 22, 675-681,1983.
10. L.J. Hornbeck,"Deformable-Mirror Spatial Light Modulators," SPIE Critical Reviews 1150, 86-102, 1989.
11. Natalie Clark, "A Silicon Eye using MEMS Micromirrors," MRS BULLETIN, 320-324, April 2001.
12. C. Quan, S.H.Wang, C.J.Tay, A.Q.Liu, H.M.Shang,"Deformation measurement of a micro-rf capacitive switch membrane using laser interferometry," Opt.Eng. 42(1) 92-97,2003.

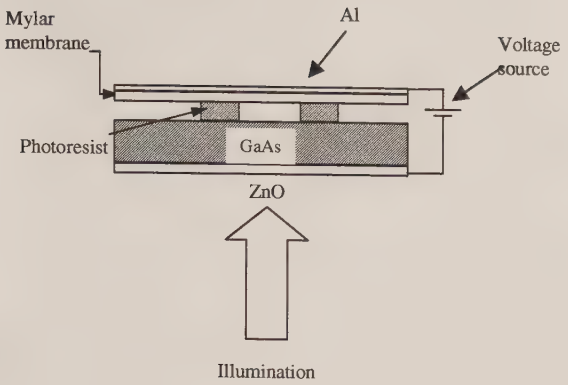


Fig. 1. Photoconductive Optically Controlled Spatial Light Modulator

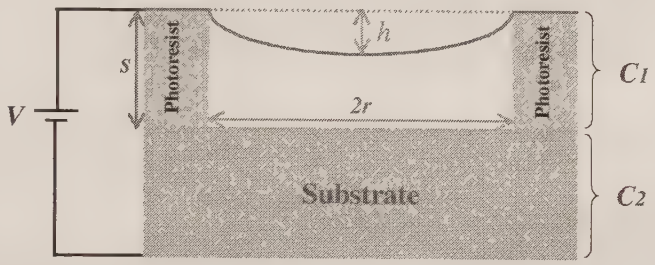


Fig. 2. The membrane driven by Dc voltage before the light illumination

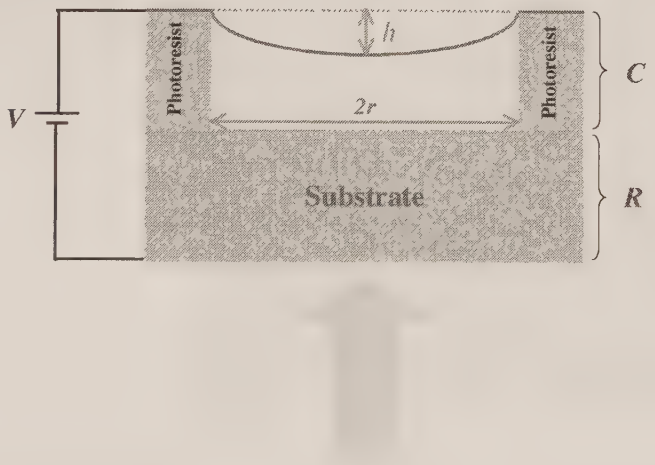


Fig. 3. The membrane driven by Dc voltage before the light illumination

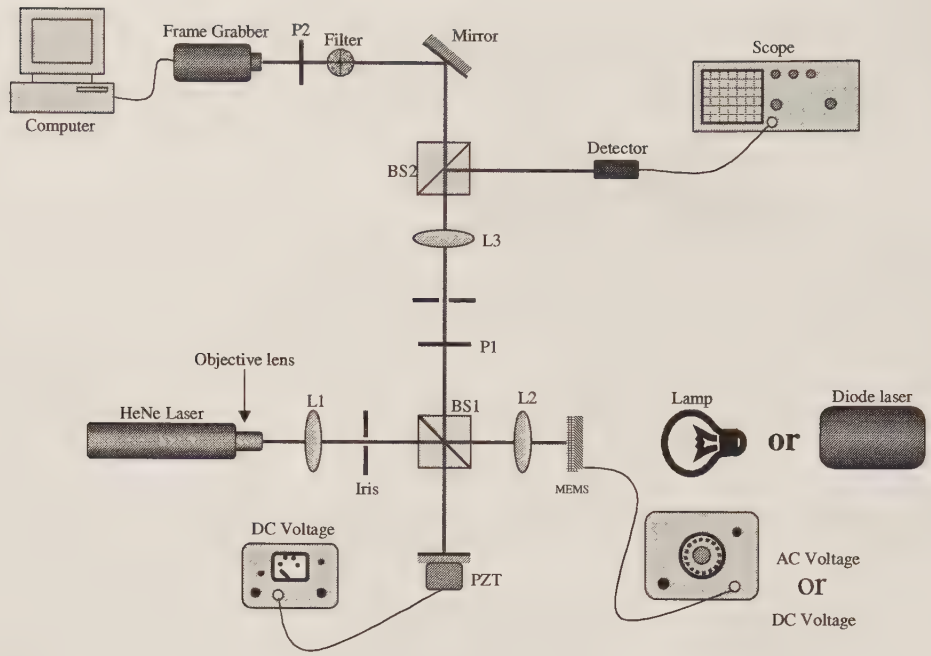


Fig. 4. Experimental arrangement

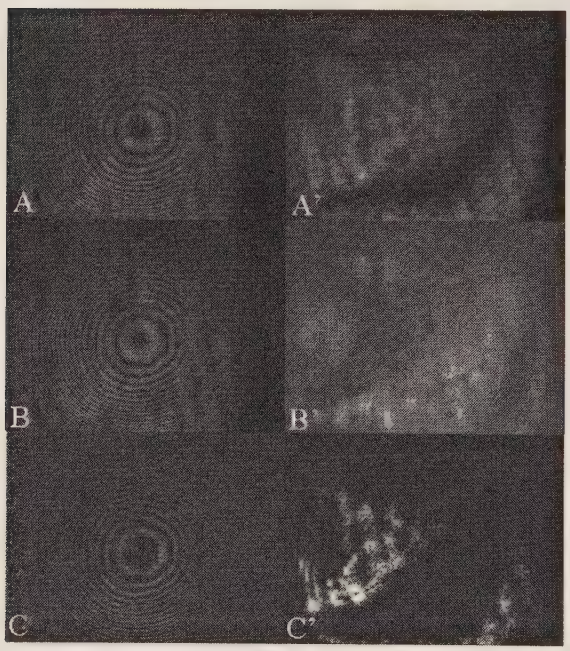


Fig. 5. (A) Fringes of punch hole sample after focusing by a lens and before lamp illumination on back; (B) Same fringes after lamp illumination; (C) The difference; (A') Fringes of punch hole sample before lamp illumination; (B') Same fringes after lamp illumination; (C') The difference

Fig. 6. (A) Fringes of 1mmx1mm sample before laser illumination on back; (B) Same fringes after laser illumination; (C) The difference; (A') Fringes of 2mmx2mm sample before lamp illumination; (B') Same fringes after lamp illumination; (C') The difference

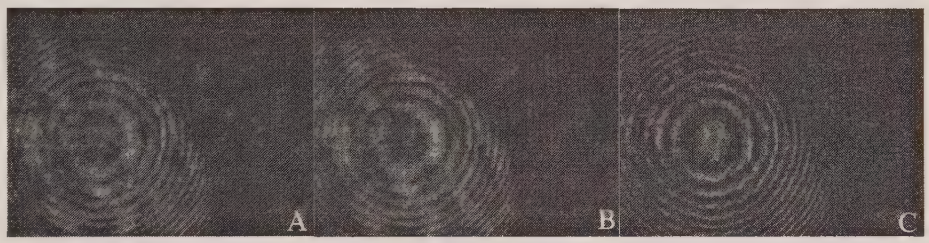


Fig. 7. (A) Fringes of 1mmx1mm sample focused on one pixel by a lens and before laser illumination on back; (B) Same fringes after laser illumination; (C) The difference.

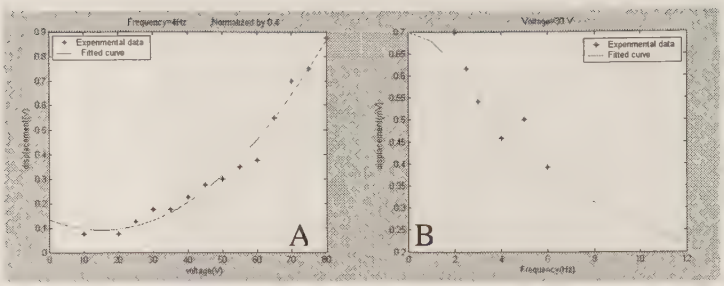


Fig. 8. (A) The deformation of membrane (1mmx1mm sample) vs. voltage at 4Hz, (B)The deformation of membrane (1mmx1mm sample) vs. frequency at 30V.

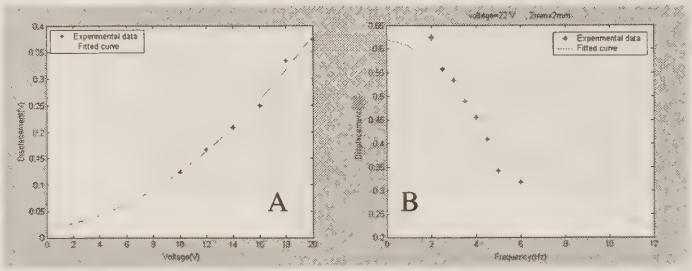


Fig. 9. (A) The deformation of membrane (2mmx2mm sample) vs. voltage at 2.5Hz, (B) The deformation of membrane (2mmx2mm sample) vs. frequency at 22V.

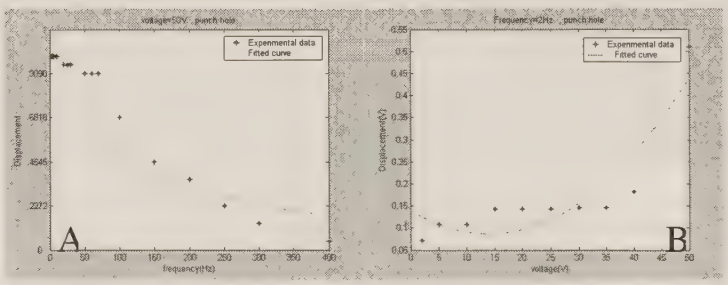


Fig. 10. (A) The deformation of membrane (punch hole sample) vs. frequency with 50Vapplied, and (B) the deformation of membrane (punch hole sample) vs. voltage for frequency 2Hz.

Optical correlator using four kilohertz analog spatial light modulators

Teresa Ewing, Steve Serati and Kipp Bauchert
Boulder Nonlinear Systems, Inc.
450 Courtney Way, #107
Lafayette, Colorado 80026

ABSTRACT

For an optical correlator system, information is encoded in four dimensions (spatially, bits per pixel and time). This paper describes an optical correlator system that was built around 8-bit, 256 x 256 spatial light modulators capable of operating at four kilohertz frame rates. The full data throughput of the optical processor is not realized due to limitations with the correlation plane detector and post processor. However, continuous operation at nearly a kilohertz was demonstrated.

Keywords: ferroelectric liquid crystal, spatial light modulator, optical correlation

1. INTRODUCTION

Optical processing has the potential to be a powerful tool, but practical limitations have prevented its usefulness. Most of the limitations are associated with the input/output (I/O) devices used in optical processors, such as spatial light modulators (SLMs) and the cameras. For an optical correlator, information is encoded in four dimensions (two spatial, one temporal and one amplitude). In the optical correlator system described here, we have improved the processing throughput by increasing the frame rate of the SLM devices. Using newly developed ferroelectric liquid crystal (FLC) materials and improved VLSI foundry processes, we have developed a 256x256 Liquid Crystal on Silicon (LCoS) SLM that provides full-bipolar amplitude modulation and operates at 4000 frames per second. The high operating speeds involved require that the SLM have good optical efficiency. Likewise, implementation in a compact correlator requires that the SLM have a small pixel pitch.

The type of modulation provided by the SLM has a large effect on the performance of the correlator. Many filtering algorithms have been developed based on different modulation schemes. The most versatile modulation is full complex-amplitude modulation, which effectively allows any filter implementation. Full complex-amplitude modulation is difficult to achieve in liquid crystal devices without combining modulation from different pixels. In most cases, the techniques used to achieve full complex-amplitude modulation make implementation impractical. With a small compromise in performance (approximately 3 dB), filter algorithms using real axis (i.e. bipolar amplitude) have been investigated and proven effective.

Most FLC materials provide polarization rotation of approximately 90°. However, the modulation depth tends to fall off with frame rate. To prevent this, new low voltage FLC materials with high spontaneous polarizations are used. Experimental results with these mixtures, however, dictated that a minimum of a 6 V field would be needed to achieve 4 kHz operating speed. With standard CMOS processes limiting the pixel voltage to 5V, the challenge was to find a high voltage solution within the confines of a commercially available silicon foundry process.

Figure 1 shows the measured response of a 2.8 μm thick FLC cell. The 1%-99% rise time (dark blue) and fall time (in pink) were measured, as well as the 10%-90% rise and fall times (in yellow and cyan). The maximum frequency achievable using the 10%-90% response times is plotted in violet. Note that at 6 V the maximum frequency is 4 KHz.

Optical Pattern Recognition XV, edited by David P. Casasent, Tien-Hsin Chao, Proc. of SPIE Vol. 5437 (SPIE, Bellingham, WA, 2004) · 0277-786X/04/$15 · doi: 10.1117/12.547404

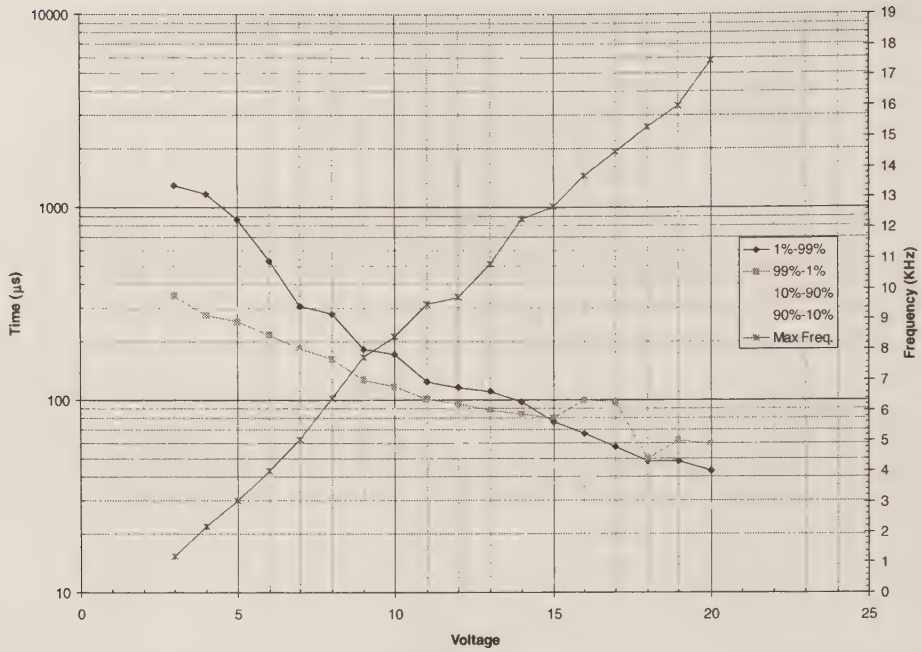


Figure 1. The measured FLC response times, and the maximum frequency at each voltage level.

2. 256x256 HIGH-SPEED SLM BACKPLANE DESIGN

Although standard CMOS processes promise 5V capability, our experience showed that there was a significant drop in voltage to the pixel pads due to the switching threshold of the transistors. Transistor switch threshold is typically 20% of the process voltage. For this reason we expected to achieve only ~4V at the pixel pads with a 5V process; not enough voltage to achieve the 4-kHz new image rate. We utilized several methods to compensate for this lack of voltage. The first method used dual complimentary transistors, one n-FET and one p-FET, which effectively eliminates the loss of voltage due to the transistor switching threshold. This technique enabled us to achieve the full 5V at the pixel pad. There are two negatives to this approach, one is the increased size of the circuitry due to the requirement for an additional transistor at each pixel, and the other is the decreased yield due to the additional circuitry. This particular foundry process could yield a pixel pitch of 10-µm or less if a single transistor design were used. However, it was felt that the increase in pixel pitch to 18-µm was acceptable given the need for the additional voltage. As it stands, the active area for the array is only 4.6 mm, and does not therefore negatively impact the size of the optical correlator.

We achieved high voltage at the pixel pad by overdriving the chip, by supplying it with 6.5V instead of 5V. This does add some risk since most silicon foundries only guarantee their process to about +10% voltage. However our past experience has shown that similar 5V chips can be driven as high as 6.8V before failure occurs. Therefore we believed that we would be able to safely run the chip at 6.5V. Unfortunately, one known negative effect of higher drive voltages is a reduced lifetime. It is not yet known how significantly the lifetime will be reduced when using a drive voltage of 6.5V. For this reason, a second SLM chip design was submitted to the foundry for production on the same wafer.

The second SLM backplane design was identical to the first chip in all aspects except for one. This second chip had a thicker gate oxide layer. The thick gate oxide layer increases the breakdown voltage of the transistors allowing for a higher drive voltage. It is known that when utilizing the thicker gate oxide layer with this foundry process that transistor voltages of 12V are achievable. However we doubted that we would be able to achieve that high of a voltage since we were not following all of the rules for the 12V process. The 12V process design rules include the use of larger transistor gates in addition to the thicker gate oxide layer. For this reason we felt that this second chip would give us more voltage capability than the first, but that it would be less than the full 12V.

The design of the 256x256 multi-level 4-KHz SLM silicon chip was optimized for a foundry process utilizing both chemical-mechanical polishing (CMP) and a "shiny" metal layer for improving the reflective efficiency of the pixels. The SLM consists of a 256x256 array of pixels that operates at a refresh rate of 24 kHz and a pixel voltage of more than 6V. The array has a pixel pitch (center to center) of 18-μm and a pixel size of 16.8-μm square. Therefore, the area fill factor is 87%.

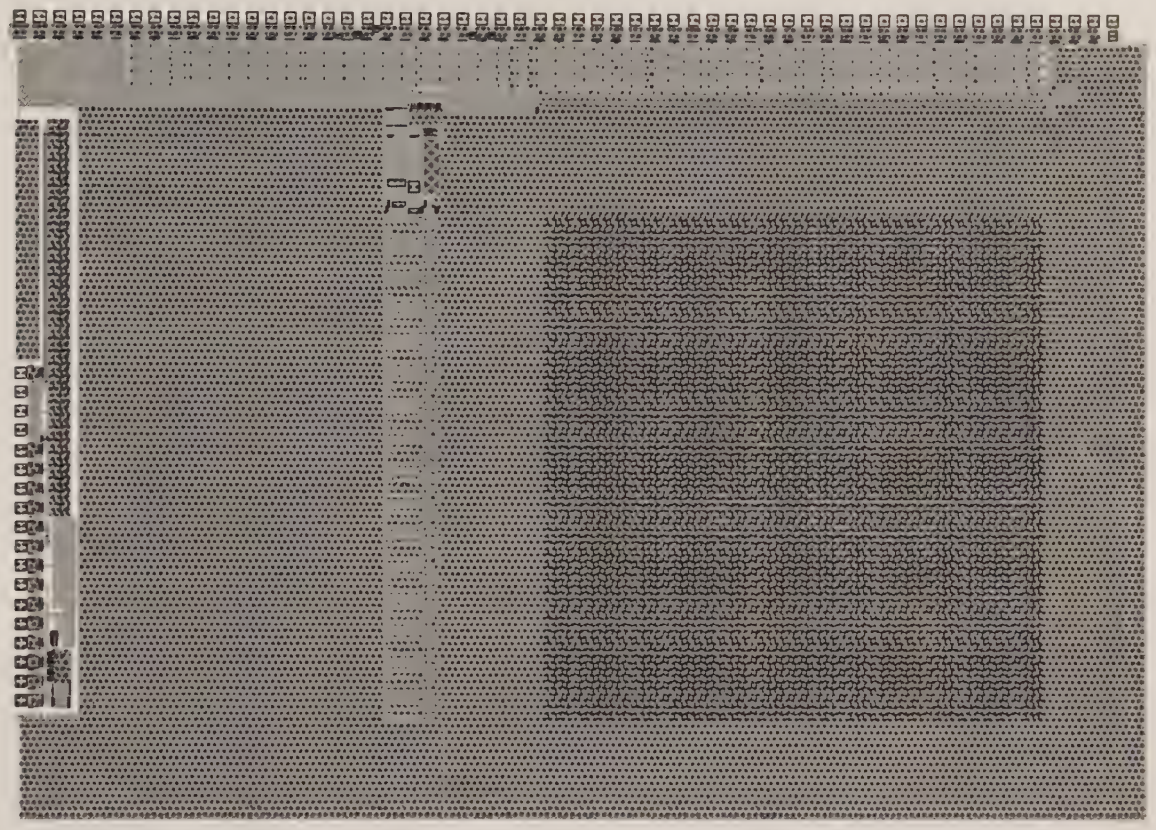


Figure 2. Drawing of the 256x256 SLM. The active area is represented by the large black rectangle. The bright purple swath along the top edge and running down the left-center of the chip represents control circuitry. The large medium purple regions covering most of the chip are "dummy" pixels. The electrical interface to the chip is constrained along the top edge. The bond pads along the left edge and the associated small circuitry region are strictly for electrical test purposes.

A new design feature included in this chip was to add a large area of "dummy" pixels surrounding the entire active area, as shown in figure 2. This was done to minimize differences in material densities across the chip. These differences in material densities can lead to a non-uniform surface flatness due to the Chemical Mechanical Polishing (CMP) steps that the foundry uses to achieve such small process sizes. Both of our current 512x512 SLMs, which do not have dummy pixels, have a slight dome because the areas outside of the active areas have a lower material density, resulting in more material being removed during the CMP steps. The use of these large "dummy" pixel regions resulted in a noticeable improvement to the overall optical flatness of the chips.

After the first SLM chips were electrically verified, they were gapped and filled. A custom flex cable was designed for the device. This flex pad is manufactured with a ceramic substrate integrated into the flexible structure, providing the SLM die a mechanically stable base surface (see figure 3). The SLM was tested by first viewing simple image patterns to verify proper pixel addressing. Once this test was successfully completed, the SLM was further analyzed for good

LC uniformity and alignment. This was evidenced by the uniformity of image patterns as they crossed the face of the SLM. Once good uniformity and alignment were demonstrated, then the SLM was further tested for such optical parameters such as switching speed, contrast ratio, and optical efficiency.

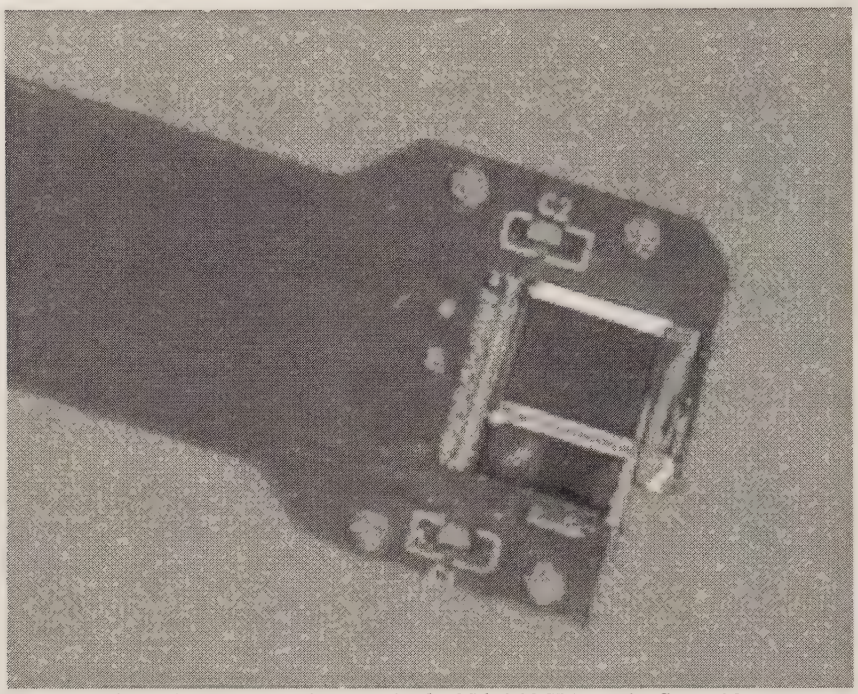


Figure 3. Photograph of a 256x256 SLM on its flex cable.

3. OPTICAL CORRELATOR DESIGN

In parallel with the VLSI design, we also developed the optics and mechanics required for a complete 4-f VanderLugt optical correlator utilizing the new 256x256 SLMs. The optical head consists of assembled and aligned optical components including two 256x256 multi-level SLMs, a DALSA 955-Hz 260x260 CCD camera, laser, passive optics, and mechanical mounts.

The basic size of an optical correlator is driven primarily by the size of the SLMs utilized in the input and filter planes. This relationship is a function of the square of the pixel pitch as shown by

$$f_{eff} = \frac{Nab}{\lambda} \tag{1}$$

where f_{eff} is the effective focal length required for the Fourier transform, N is the number of pixels, a is the pixel pitch for the first element, b is the pixel pitch for the second element, and λ is the wavelength of the laser source. For a typical 4-f optical correlator, the total optical path length is approximately four focal lengths.

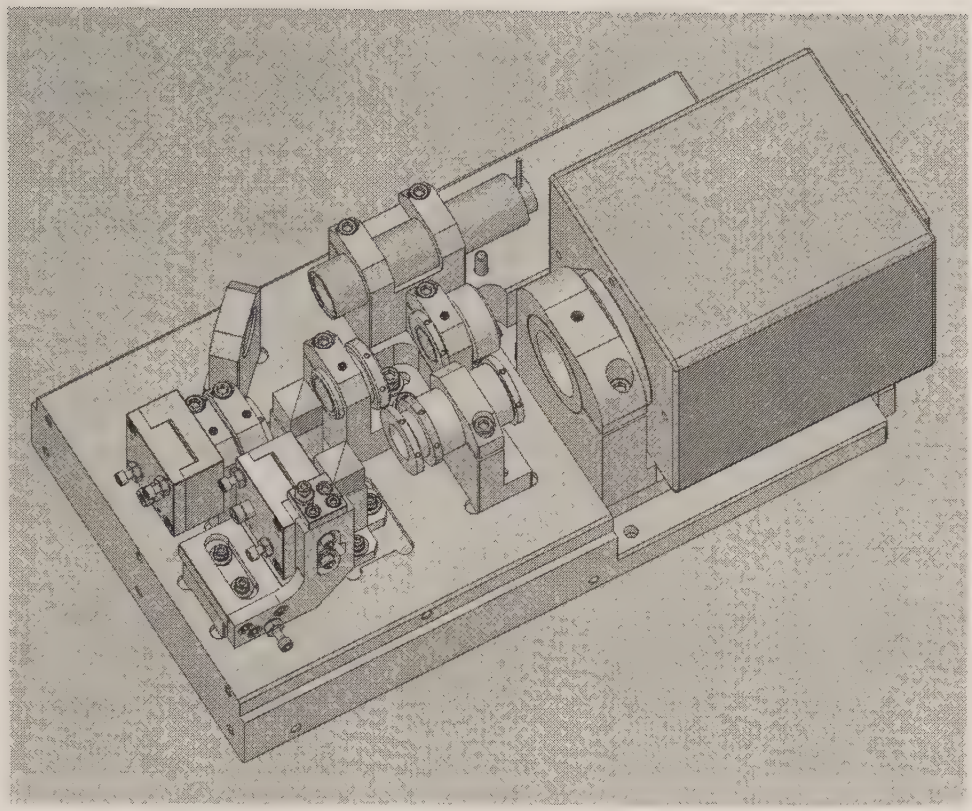


Figure 4. Upper left view of Optical Head less Cover, Panels and Frame.

The optical layout was designed to give the smallest possible optical correlator head. This 4-f correlator was designed with a 256x256 SLM pixel pitch of 18-μm, a DALSA camera pixel pitch of 10 μm, and a 680-nm laser diode. This resulted in a total optical path length of approximately 380 mm or about 15 inches. This made for a fairly small correlator, in which the size of the package was primarily driven by the large size of the DALSA camera. The camera body is quite large at 3.5" x 3.5" x 4.1". No smaller camera with the required speed and format was commercially available.

Figure 5 shows the final layout of the optical correlator. Polarized light exits the laser, is reflected by Mirror 1 to Beam splitter 1, then directed through a half-wave plate, Retarder 1, and onto the Input SLM. The halfwave plate is used to rotate the polarization exiting the beamcube into the optical axis orientation of the SLM. The beam is then reflected from the Input SLM back through Retarder 1 to Beam Splitter 1. In this case, the half-wave retarder is acting to translate from the SLM optical orientation into the beamcube frame of reference. Beam Splitter 1 then transmits the beam to Lens 1, through a quarter-wave plate, Retarder 2, and onto Mirror 2. Mirror 2 then reflects the beam back through Retarder 2 and Lens 1 to Beam Splitter 1. The double pass through Lens 1 Fourier transforms the input image. The double pass through the quarter wave plates rotates the polarization by 90 degrees as required by the beam splitter. Beam Splitter 1 then reflects the beam to Beam Splitter 2, where it is directed through a half-wave plate, Retarder 3, and onto the Filter SLM. Retarder 3 aligns the polarization exiting the beamcube into the SLM's frame of reference. The beam is then reflected from the Filter SLM back through Retarder 3 to Beam Splitter 2. The beam then passes through Lens 2 and Lens 3 and is focused onto the detector array of the Camera. The Len2/Len3 pair performs the second Fourier transform.

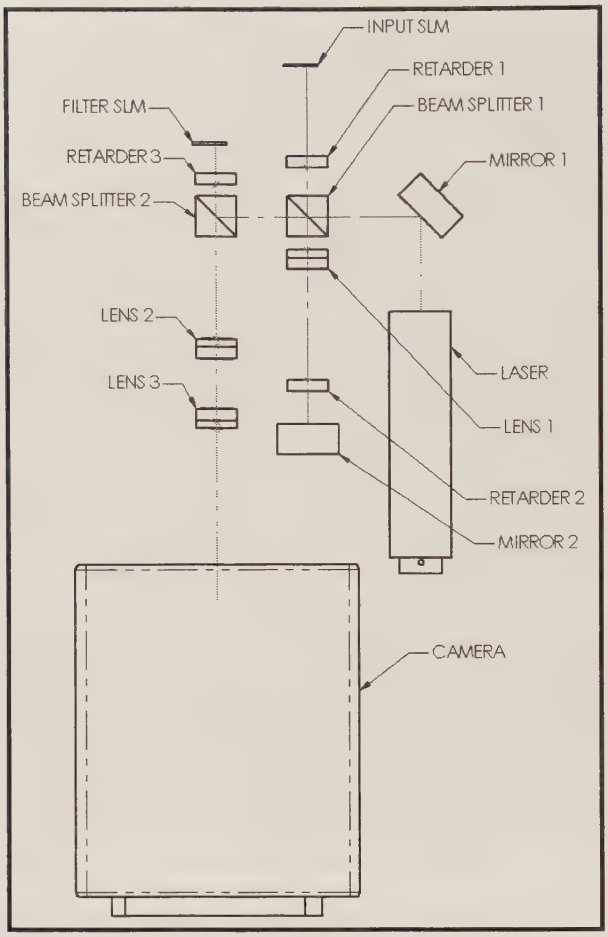


Figure 5. Optical layout of the Optical Head (top view).

The laser chosen for this correlator system is manufactured by MicroLaser Systems, part number SLM-9C-685-12. The center wavelength of this device is 685 nm. The output power is 15 mW, and has a beamwidth of 9 mm. The Camera is a DALSTAR Series Camera, part number CA-D6-0256, manufactured by DALSA, Inc. This camera has 260x260 square pixels on 10 μm centers. The camera has a maximum frame rate of 955 frames/per second. To achieve this data rate, data is sent out of the camera on four separate data cables. An electronic shutter is designed into the camera, and is integral to timing of the correlator.

4. RESULTS

4.1. Test of Thick Gate Oxide Chip

In the voltage test, the performance of the two chips (low voltage and thick gate oxide (high voltage)) was directly compared. Both chips were operated at their normal clock rate while the supply voltage was adjusted. The current into the chips was monitored as the voltage was increased until the chips electrically failed. This failure mode was correctable by lowering the voltage, so the chips were not tested to destruction. The high-voltage (HV) version of the chip ran at a considerably lower current rate. If it is assumed that heating due to power dissipation limits chip lifetime, then the HV version is capable of operating at a higher voltage. At 5.5 volts, which is the maximum operating voltage specified by the foundry for the VLSI process, the low-voltage (LV) version of the chip consumes approximately 11 watts. For the HV version, the supply voltage increased to nearly 7 volts before 11 watts of power was consumed by the chip. By increasing the supply voltage further, the LV chip electrically failed at approximately 9 volts and the HV chip failed at 10.2 volts. These data points suggest that the additional oxide and larger gate widths have increased the

maximum operating voltage of the chip by over one volt. Therefore, these tests suggest that the HV version is capable of operating safely at 6.5 volts to 7 volts. A 6.5 to 7 volt operating range satisfies the LC material requirement since an operating voltage of 6 volts provides 4 KHz switching rates. Therefore the HV version of the 256x256 chip with the thick gate oxide was selected for the 4 KHz SLMs.

4.2. Maximum Pixel Pad Voltage Achieved

The pixels on the 256x256 chip contain both p-FET and n-FET transistors allowing full rail-to-rail sampling of the inputs so that the maximum pixel pad voltage can be equal to the supply voltage to the chip. We could achieve a 7V pixel pad voltage if we supplied 7V to the 256x256 (HV version) chip. However, the final 4 kHz SLMs, have a supply voltage of 6.5V and a data (pixel) voltage of 0 to 6.25V. These voltages meet the drive requirements of the liquid crystal and are within the safe operating range of the chip.

4.3. Response Time Measurements

Other tests performed on the 4 kHz SLMs include intensity and gray-scale tests at multiple speeds as well as contrast ratio measurements. The throughput for an FLC SLM usually decreases as speed increases. The opposite is true of gray-scale; it increases as the speed increases.

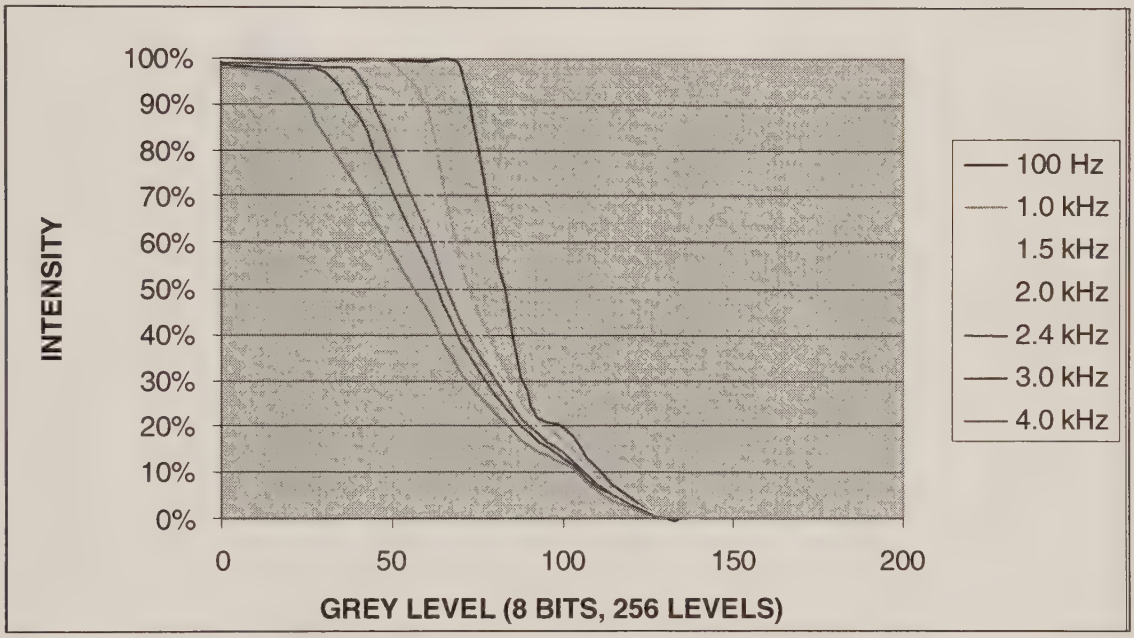


Figure 6. Grey Level vs. Intensity/Throughput plots for various speeds.

Figure 6 shows the results of the gray-scale and intensity tests at various speeds. The 4 kHz curve has the best gray-scale performance out of all the speeds measured. It does however have a slight loss of peak intensity of about 3%. The contrast ratio for these devices was also measured. The contrast ratio is determined by maximizing the OFF state for the SLM and then measuring the ratio of the ON and OFF states. The contrast ratio measured for these 4 kHz devices is 400:1.

4.4. Diffraction Efficiency Tests

In this test, the optical power of a 632nm HeNe laser was compared with the power diffracted into the zero-order by the 256x256 VLSI backplane. The zero-order diffraction efficiency of the raw VLSI backplane was found to be 61.0%. The same tests were again repeated, but with a 256x256 SLM filled with liquid crystal. The zero-order diffraction efficiency for the SLM in this case was found to be 54%.

4.5. Device Flatness

The new backplanes were tested for flatness using an interferometer. Two different die were tested, one from the edge of the wafer, and one from the center of the wafer. The center die had a somewhat cylindrical shape, and a peak-to-valley variation of 0.384 waves (@ 633 nm). The edge die had a severe wave shape, but a lower peak-to-valley variation of 0.276 waves.

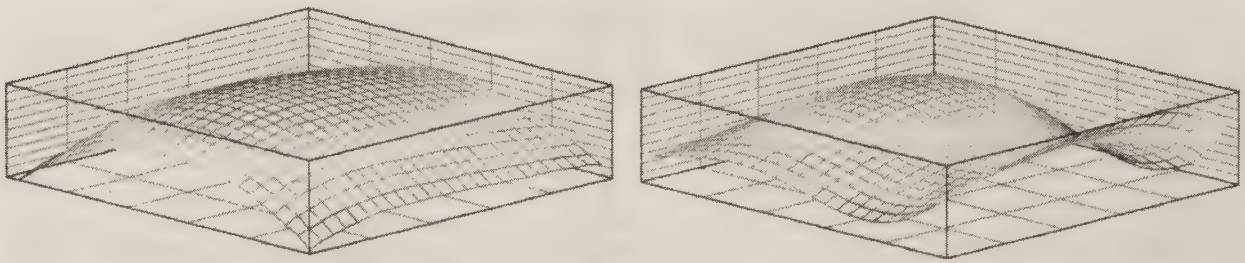


Figure 7. Flatness measurements of a center die (left) and an edge die (right).

4.6. Verification of Successful Chip Imaging

The 4 KHz 256x256 (HV) SLMs were built and then tested to evaluate their performance. In the first test of the 4 KHz SLMs, we downloaded various images to the SLMs to verify correct loading of the data as well as to evaluate the optical quality of the SLMs. Sample imagery from the 4 kHz SLMs used in the correlator is shown in Figure 8. The images verify that data is loading correctly on the SLMs. The images also demonstrate the quality of contrast and gray-scale of these 4 kHz devices.

Figure 8. Imagery displayed on the new 4kHz 256x256 SLMs

4.7. Correlation Results

Once the system was integrated and the correlator optical path was properly aligned, a number of test images were run on the system. The images below show the correlation spot achieved when a 8x8 pixel square was placed at various x-y locations. In each pair, the input pattern is shown on the left and the correlation plane result is shown on the right.

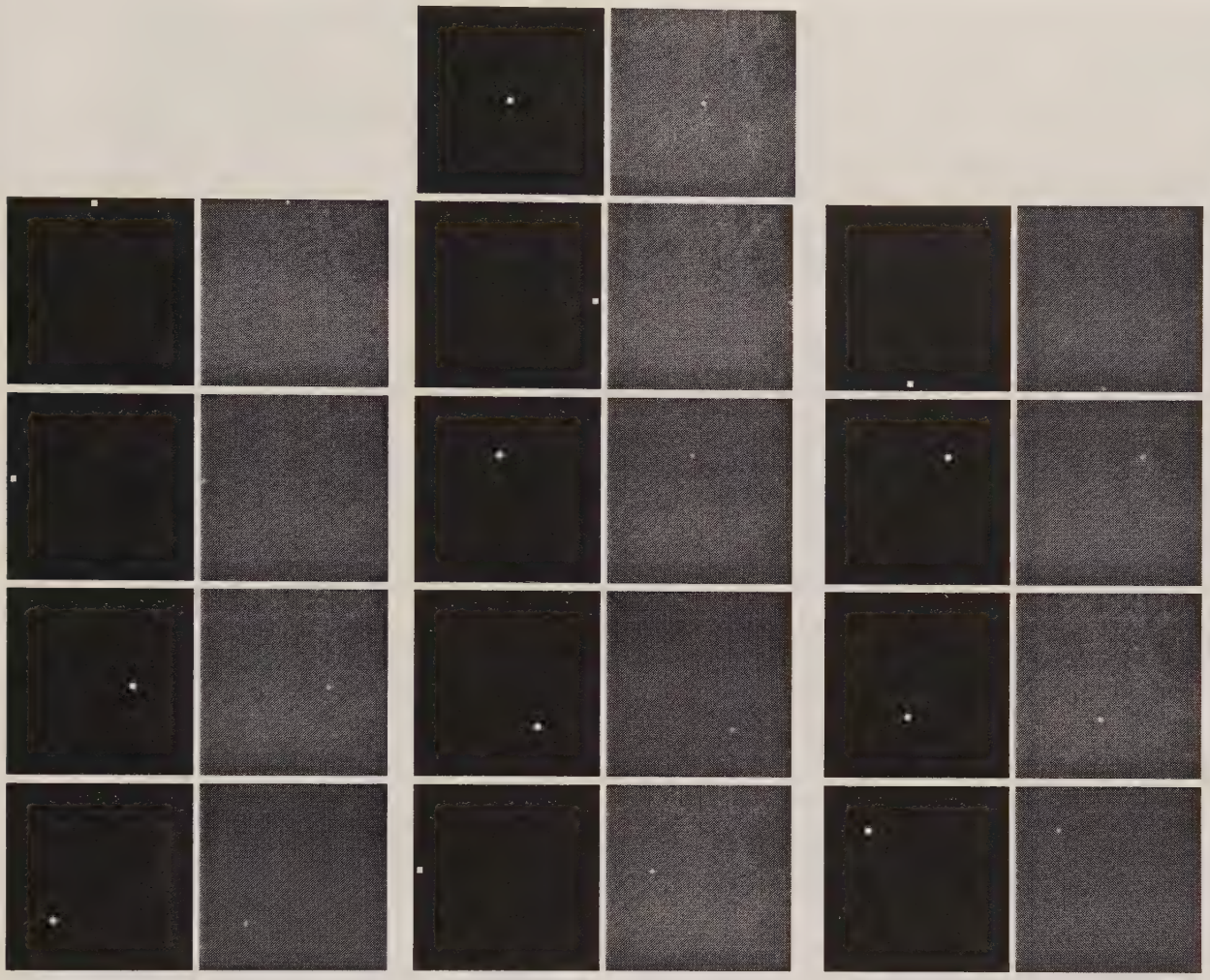


Figure 9. Scale invariant correlations at different x-y locations.

Correlation results with scaling of the input image size is shown in the images below. In this test, the size of the X/O pattern is increased, as well as the size of the "X" filter image. In each pair, two bright correlation peaks, corresponding to the "X" hits are shown. In addition, faint cross-correlations can be seen in the "O" patterns.

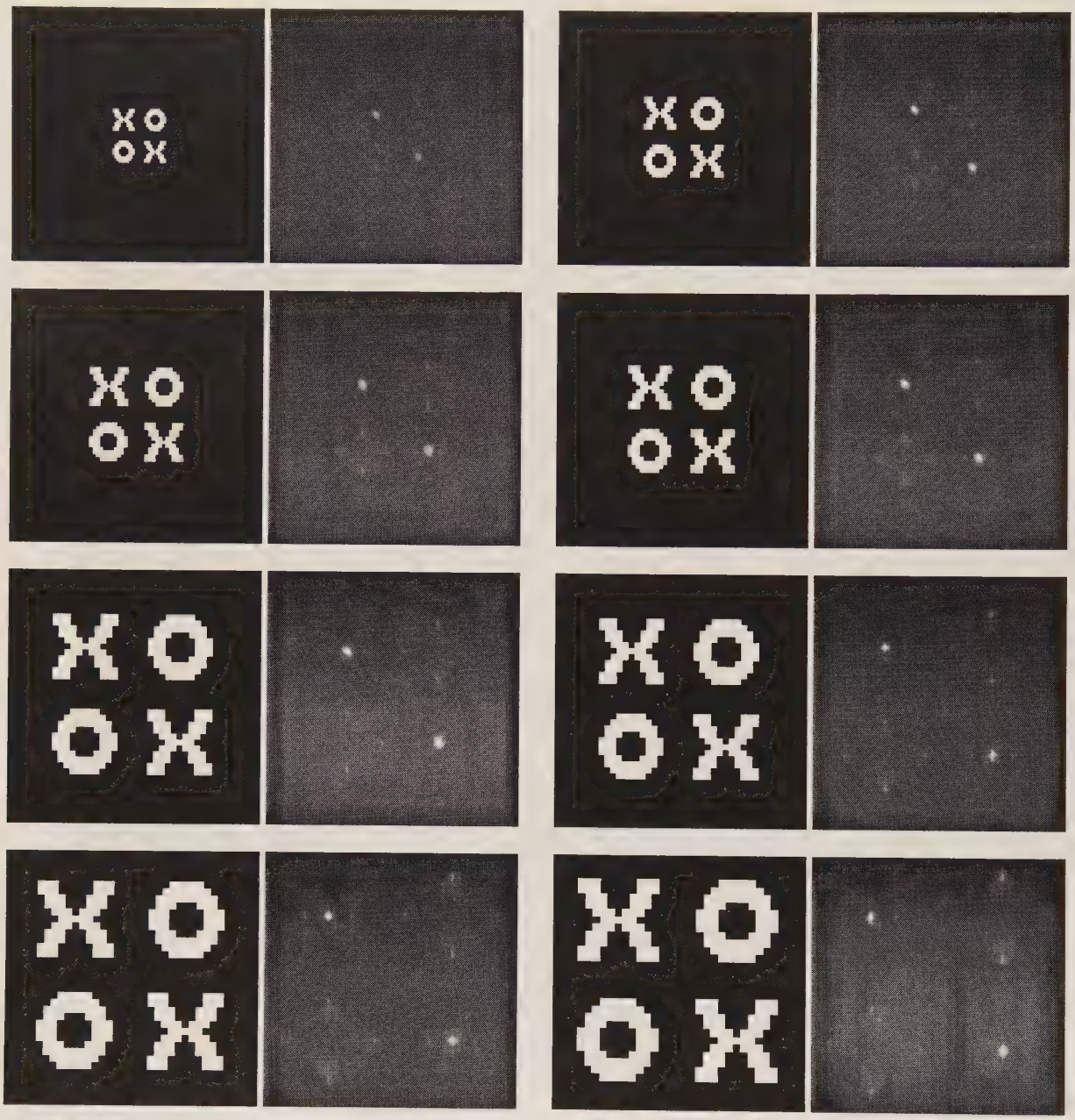


Figure 10. Correlation results with feature size scaling.

5. CONCLUSIONS

Under this program, BNS has successfully produced the first 4 kHz capable optical correlator. The actual output of the unit is limited by the output capacity of the Dalsa camera to 979 frames per second. However, the rest of the correlator is capable of 4,000 frames per second.

After reviewing the results presented, it is clear that 4 kHz operation in a 256x256 correlator is achievable. Although there are some issues to be worked out, we are optimistic about the possibility of scaling the resolution of the correlator to 512x512, while maintaining 4kHz operation. We believe that the data rate to the SLM is at the maximum speed achievable with commercially available circuitry. Therefore to scale to 512x512, the width of the data bus to the SLM will need to be increased. We predict that we will be able to maintain the current pixel pitch (18 μm) in a 512x512 SLM VLSI design. This will result in an active array area of 9.2mm x 9.2mm. Although the maximum die size is limited to the foundry's reticule size, there would be sufficient real estate to allow for dummy pixels, which have proven very effective in this program at reducing curvature of the VLSI backplane.

Once again, the camera will be the limiting factor to the operational speed of the optical correlator system. Currently available 512x512 area scan cameras are limited to 200-300 frames per second. Camera manufacturers do, however, offer high-speed CMOS area array sensors (without electronics) that would be of much higher speeds (640x480, 1300 frames per second).

ACKNOWLEDGEMENTS

The authors would like to thank Rick Martin, Anna Linnenberger, Jim Brooks, Justin Knowles, and Owais Mughal for their assistance with this project. In addition, we gratefully acknowledge the funding and support of Jim Kirsch and the U.S. Army Aviation and Missile Command under contract #DAAH01-01-C-R172.

REFERENCES

1. D. McKnight, K. Johnson and R. Serati, "Smart spatial light modulators using liquid crystals on silicon," *Quant. Elec.* **29**(2) pp 699-714, 1993.
2. S. Serati, G. Sharp, R. Serati, D. Mcknight and J. Stockley, "128x128 analog liquid crystal spatial light modulator," *Proc. SPIE*, **2490**, pp. 378-387, 1995.
3. K. Bauchert, S. Serati, G. Sharp and D. McKnight, "Complex Phase/Amplitude Spatial Light Modulator Advances and Use in a Multispectral Optical Correlator," *Proc. SPIE*, **3073**, pp 170-177, 1997.
4. S. Serati and K. Bauchert, "Sampling technique for achieving full unit-circle coverage using a real-axis spatial light modulator," *Proc. SPIE,* 3715, pp 112-119, 1999.
5. P.M. Birch, R.C.D. Young, D.M. Budgett, C.R. Chatwin, "Fully Complex Optical Modulation using an Analogue Ferroelectric Spatial Light Modulator" *Optics Communications*, **175,** pp 347-352, (2000).

Distributed multi-sensor processing, decision making, and control under constrained resources for remote health and environmental monitoring

A. Talukder, PhD
CHLA/Univ. of Southern California & Jet Propulsion Laboratory/NASA
Los Angeles, CA
Email: Atalukder@chla.usc.edu

Tanwir Sheikh
CHLA/USC
4650 Sunset Blvd., MS 88
Los Angeles, CA 90027

Lavanya Chandramouli
CHLA/USC
4650 Sunset Blvd., MS 88
Los Angeles, CA 90027

Abstract

Previous field-deployable distributed sensing systems for health/biomedical applications and environmental sensing have been designed for data collection and data transmission at pre-set intervals, rather than for on-board processing These previous sensing systems lack autonomous capabilities, and have limited lifespans. We propose the use of an integrated machine learning architecture, with automated planning-scheduling and resource management capabilities that can be used for a variety of autonomous sensing applications with very limited computing, power, and bandwidth resources. We lay out general solutions for efficient processing in a multi-tiered (three-tier) machine learning framework that is suited for remote, mobile sensing systems. Novel dimensionality reduction techniques that are designed for classification are used to compress each individual sensor data and pass only relevant information to the mobile multisensor fusion module (second-tier). Statistical classifiers that are capable of handling missing/partial sensory data due to sensor failure or power loss are used to detect critical events and pass the information to the third tier (central server) for manual analysis and/or analysis by advanced pattern recognition techniques. Genetic optimisation algorithms are used to control the system in the presence of dynamic events, and also ensure that system requirements (i.e. minimum life of the system) are met. This tight integration of control optimisation and machine learning algorithms results in a highly efficient sensor network with intelligent decision making capabilities. The applicability of our technology in remote health monitoring and environmental monitoring is shown. Other uses of our solution are also discussed.

Keywords: Biomedicine, distributed sensing, environment monitoring, evolutionary computing, genetic algorithms, control, optimisation, mobile health monitoring, machine learning, pattern recognition, resource management, sensor fusion, sensor networks.

1. INTRODUCTION

While research has advanced the state of low-powered devices, and wireless technology has made possible the transmission of remote data, the ability to intelligently process such data in-situ (on-board), adapt to changing environments, and take corrective action in response to events/anomalies has been very limited. Few examples that have addressed mobile sensing from a hardware and infrastructure development perspective has been the Intel Proactive Health Strategic Research Project headed by Eric Dishman, the JPL/NASA sensor networks initiative [sensorwebs.jpl.nasa.gov], the Intel EcoSense and Heterogeneous Sensor Networks projects headed by Lakshman Krishnamurthy [1], and the UC Berkeley SmartDUST project [2].

The primary hurdle to intelligent mobile sensing has been the lack of adequate distributed machine learning algorithms that can run on devices with limited computing capability, and lack of resource management algorithms designed specifically for sensor networks. Current machine learning algorithms, such as traditional neural networks typically are slow and require lots of computing power, and large amounts of memory, and also require lots of training data, and therefore perform inadequately with high-dimensional inputs [4]. Additionally, current statistical multisensor fusion algorithms [3] cannot handle missing sensor data in an adequate manner, and nor can they handle distributed processing in low-bandwidth conditions.

The problem of controlling the operations in a sensing system while keeping desired goals in mind can be categorized into any of several domains, such as resource allocation, optimisation, or adaptive control.

Optical Pattern Recognition XV, edited by David P. Casasent, Tien-Hsin Chao, Proc. of SPIE Vol. 5437 (SPIE, Bellingham, WA, 2004) · 0277-786X/04/$15 · doi: 10.1117/12.548080

Solutions to these set of similar problems range from market-inspired economic models using auctioning techniques [9], feedback control mechanisms to real-time scheduling [10], the spring scheduling algorithm that can gaurantee tasks via on-line admission control [11], among others. The problem with the spring algorithm is that it is a "open-loop" algorithm. Its schedules cannot be adjusted dynamically and therefore fail in uncertain environments. Evolutionary algorithms are attractive for use in complex, dynamic problems where new events

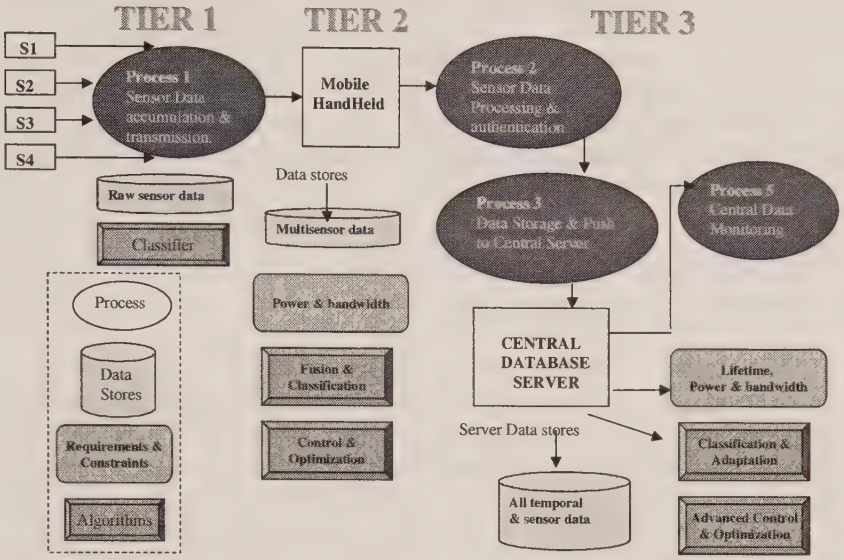


Figure 1: Mobile distributed sensing concept

change the system and environmental settings frequently, and have been noted [12] to outperform traditional controls and optimisation-based solutions in test scenarios. We propose the use of Genetic algorithms for real-time control of the sensor network system. Genetic algorithms have been shown to perform at least as well as other deterministic optimisation techniques, and even outperform competing approaches in dynamic scheduling and control problems. In addition, they do not require computation of a gradient and/or inversion of complex, large matrices, which makes them suitable for applications with limited computing power. The iterative nature of the GA techniques, of course, could limit their use in real-time optimisation, but we discuss methods to combine off-line design optimisation methods with on-line dynamic optimisation to facilitate their use in adaptive control in sensor networks.

We present a new paradigm for **distributed remote sensing in general,** where novel feature extraction (dimensionality reduction) algorithms are used to transmit sensory data (Tier 1 in Fig. 1) in low bandwidth while retaining critical classification information. A new mixture of experts architecture that can handle missing partial sensory data is used (Tier 2 in Fig. 1) classify events based on the extracted low-dimensional features. An efficient adaptive control technique using evolutionary algorithms is used in Tier 2 (Fig. 1) to control the system under changing environments. Our intelligent mobile sensing architecture is shown in Fig. 1. We show the feasibility of our approach on a mobile health monitoring example (for soldiers, patients, and alcoholics), and for greenhouse environmental monitoring, distributed sensing, and control at the same time. Both applications are heavily dependent on low-powered computing platforms, multisensor fusion, and low bandwidth and have lots of similarities. Therefore our solution is applicable to a variety of domains where real-time distributed sensing is required.

In Section 2, we discuss our dimensionality reduction algorithms and compare them theoretically with other dimensionality reduction techniques. This is followed by a detailed description of our mixture of experts architecture in Section 3. Our adaptive control algorithms for sensor networks using evolutionary techniques are detailed in Section 4. Results on simulated data are presented in Section 5.

2. DIMENSIONALITY REDUCTION FOR WIRELESS SENSING

Dimensionality reduction of sensory data is an important part of our sensing system for low-bandwidth conditions. While features can reduce the load of data transmitted by two or more **orders of magnitude**, they

should retain useful information for detecting important events. We use nonlinear features that incorporate higher order correlations in the input data, but can be computed without any overhead compared to the computation of linear features. This is a very novel aspect of our proposed research. Conventional feature extractor and discriminant function methods such as Fisher and it's variations such as the optimal discriminant function [3], Fukunaga Koontz, etc., separate the means of projections per class to achieve discrimination. We have shown [4] that this is not optimum and that our features provide better results than such standard methods on various applications

We now highlight our dimensionality reduction algorithm. Without loss in generality, we consider 2 classes $\{\mathbf{x}_{1p}\}$ and $\{\mathbf{x}_{2q}\}$ with P and Q samples, with means $\underline{\mu}_1$ and $\underline{\mu}_2$, and with covariance matrices $\underline{C}_1$ and $\underline{C}_2$. We determine M orthogonal transformations $\underline{\varphi}_m$ that produce M new class 1 and class 2 features $y_{m1p} = \underline{\varphi}_m^T \underline{x}_{1p}$ and $y_{m2q} = \underline{\varphi}_m^T \underline{x}_{2q}$ in a reduced space (M< P and Q) to simplify calculations. To determine $\underline{\varphi}_m$, we require that $\underline{\varphi}_m^T(\underline{C}_1+\underline{C}_2)\underline{\varphi}_m$ be minimized for all $\underline{\varphi}_m$ (this minimizes the spread of the output features for each class, and is conventional). We also require the $\underline{\phi}_m$ to maximize $\left(y_{m1s} - y_{m2q}\right)^2$ for all samples in classes 1 and 2 *(this is very different from conventional methods that only maximize the separation of the means of the projections per class)*. This requires maximization of $\underline{\phi}_m^T \underline{R}_{12} \underline{\phi}_m$, where $\underline{R}_{12} = \sum_s \sum_q (\underline{x}_{1p} - \underline{x}_{2q})(\underline{x}_{1p} - \underline{x}_{2q})^T$ is a vector outer product difference matrix. *This step differs considerably from all prior methods.* Thus, we calculate the $\underline{\phi}_m$ and associated features to maximize $E_D = \sum_m (\underline{\phi}_m^T \underline{R}_{12} \underline{\phi}_m) / (\underline{\phi}_m^T (\underline{C}_1 + \underline{C}_2) \underline{\phi}_m)$. This E_D is our *new discrimination measure* that makes our features preferable to standard ones. The solution for the $\underline{\phi}_m$ are [13] the dominant eigenvectors of

$$\left[\underline{C}_1 + \underline{C}_2\right]^{-1} \underline{R}_{12} \quad . \qquad (1)$$

This yields *linear discriminating features*. (LDFs). Extensions to more than 2 classes are obvious [13].

The new E_D performance measure solution in Eq. (1) handles cases when multiple clusters are present in any class. In such cases, the distributions of the data per class are asymmetric and nonlinear features are needed. The linear DF algorithm easily extends to the case of nonlinear DF features using the new algorithm now described. We write the general quadratic description of the input $\underline{x}$ data in an extended higher-order H

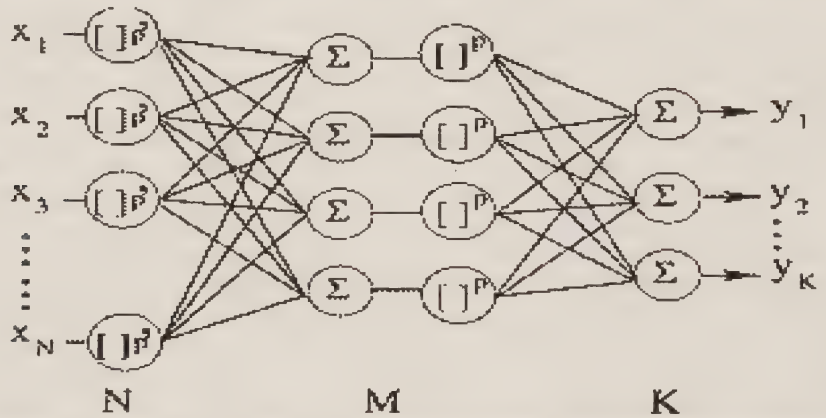


Figure 2: Dimensionality reduction for classification implemented as a neural network with closed-form solutions

space as

$$\underline{x}_H = \left[x_1 \quad x_2 \quad \cdots \quad x_N \quad x_1^2 \quad x_1 x_2 \quad \cdots \quad x_N^2\right], \quad (2)$$

where $H = N + N(N+1)/2$ is the dimension of what we refer to as a higher dimensional (quadratic) $\underline{x}_H$ version of $\underline{x}$. It contains each input (raw sensor data) raised to the power 2 as well as all cross-terms (e.g. $x_1 x_2$). We produce associated covariance ($\underline{C}_{1H}$ and $\underline{C}_{2H}$) and vector outer product difference ($\underline{R}_{12H}$) matrices in this higher-order input space. The NLDF (nonlinear DF) solution for these nonlinear features (the

$\underline{\phi}_{mH}$) are similar to (1), they are the dominant eigenvectors of $[\underline{C}_{1H} + \underline{C}_{2H}]^{-1} \underline{R}_{12H}$. This is a quite new approach to solving nonlinear problems in which the *input feature space problem is linear in the extended (H) space in which the problem is solved, but is nonlinear in the actual input space used.* The solutions are available in *closed form.* They are thus preferable to neural net and other iterative solutions.

We now address a new algorithm that we use for the present high-dimensional sensor data case where N >> S (typically we expect small training set size S and large image input size N). In this case, the $\underline{C}$ estimates are poor and $\underline{C}^{-1}$ cannot be computed. We propose a new 2-stage algorithm as we now detail. It is now vital to reduce the number of features M used; i.e. M<S is essential, else $\underline{C}^{-1}$ cannot be computed as in (1) or its higher-order version. At each stage, we consider nonlinear polynomial transformations of the data without cross product terms. The first stage involves a simpler $E'_D = \underline{\phi}_m^T \underline{R}_{12} \underline{\phi}_m$ measure to optimize that does not require $\underline{C}$ to be inverted. We raise each input sample to the power p' and thus produce a new vector $\underline{x}_{H1} = \left[x_1^{p'} \; x_2^{p'} \ldots x_N^{p'} \right]$. In the second stage we take these new y' features and raise each to a power p to produce new input vectors $\underline{x}'_{H2} = \left[y_1^p \ldots y_K^p \right]^T$. In this new {y} space with reduced dimensionality K, we then solve for the final $M \leq K$ features $\underline{y}$ that maximize E_D.

We refer to these as nonliner maximum representation and discrimination features (NL MRDF) . These are quite novel as the solution is obtained in closed form; they are thus expected to be superior to iterative NL feature solutions such as NL PCA and Support Vector Machines [4]. These are the features we will use as inputs to the multisensor fusion algorithms. They can be realized by a novel neural net with adaptive activation functions (these implement raising the features to the powers p and p'). The neural network implementation of our dimensionality reduction technique is shown in Fig. 2. Our algorithm has been applied to many spatial (image) datasets for inspection of nuts, for machine parts recognition, and to faces [5]. It was seen to outperform standard discriminant functions and PCA features In all cases, initial results were excellent and our algorithm performed better that others. Space does not permit extensive examples to be included (they are in the references noted).

3. MULTISENSOR FUSION AND CLASSIFICATION

The classification scheme for distributed multiple sensors should be capable of handling missing sensory data or highly degraded sensor data to the point where information from a sensor is no longer considered reliable. We propose a novel classification architecture that uses the features from each sensor in making a decision while robustly **handling missing or degraded sensor data**. We consider the decision making process as a soft competition scheme where multiple classifiers are trained on "incomplete" information (or non-optimal features from the multiple sensors); the results from each classifier is then combined using a mixture model to make a final decision.

Ideally, if all information about the observed process (OP) were available, an optimal feature set would yield the maximal classification performance. This would correspond to a classification process $y = g(f_{opt}(OP))$. $f_{opt}(OP)$ corresponds to an optimal feature set extracted from the true observed process (OP), and g( ) corresponds to a nonlinear mapping of the classifier using the optimal feature set. However, in truth, multiple sensors are used to capture only partial information about the OP. Furthermore, features are extracted from each sensor that further reduces the information extracted from the OP.

One solution to the classification problem from the feature sets from each sensor is to pool all the features together and classify the OP with the combined feature set. However, this may result in a large parameter set for the classifier if the number of sensors are large. The curse of dimensionality in many machine learning algorithms (as noted in [4]) therefore limits the possibility of pooling all sensory features into a single "macro-feature". Therefore, we view the classification process as a hierarchical mixture of experts, where each expert considers only one feature set from a single sensor. While mixtures of experts [8] are popular in speech recognition and image/text classification, they are typically not designed to handle multisensor feature data and most importantly cannot handle missing/partial input data. In our linear combination scheme, we adopt a soft competition mechanism on different feature sets to determine weights in an optimal way for linear combination. An expectation-maximization (EM) [6] learning algorithm is also proposed for parameter estimation in the

linear combination scheme. Due to limited space, we briefly outline our classification scheme to detect events of interest in the OP.

3.1. Classification of Individual Sensory Data

We use a statistical mixture of Gaussians classifier [7] to classify each sensor data from their low-dimensional features. We use an Expectation Maximization [6,7] approach to determine the parameters of the Gaussian Mixture (mean, variance, and mixture coefficients).

Since the MRDF features are mutually uncorrelated, the covariance matrix of the Gaussians is a diagonal matrix (off-diagonal elements are zero). This decorrelation of the input feature data eases the parameter estimation problem tremendously, and is critical in cases where training data is limited.

3.2. Multi sensor Classification

We consider the task of pattern classification on different feature sets, each from a different sensor. We develop a general paradigm where R different classifiers are used per sensor set, and S different sensors are used to observe the environment. Thus, a total of S*R classifiers are used in decision-making; each classifier C_{sr} takes feature set S (from sensor S) as input and yields a probability (or confidence) vector of the sample to belong to each of the M classes, that satisfy the following conditions: $P_{sr}(x_{sr}) = [P_{sr}^{1}(x_s) \ldots\ldots P_{sr}^{m}(x_s) \ldots P_{sr}^{M}(x_s)]$; $\Sigma_{m=1..M} P_{sr}^{m}(x_s) = 1$. We assume that there are priors $\Gamma = \{\rho_{sr}, \theta_s\}$ associated with each individual classifier. For a general input observation O, with corresponding sensor readouts $\{x_1, \ldots x_S\}$, the posterior probability of O belonging to class *m* is $P(y = m \mid O, \Gamma) = \Sigma_{s=1..S} \Sigma_{r=1..R} \rho_{sr} \theta_s P_{sr}^{m}(x_s)$, where $\Sigma_{r=1..R} \rho_{sr} = 1$ and $\Sigma_{s=1..S} \theta_s = 1$. Intuitively, θ_s corresponds to the importance of the features extracted from sensor s, and ρ_{sr} corresponds to the importance of classifier *r* for the sensor *s* output features. Therefore, the θ_s play the role of soft competition on the S different sensors (or their feature sets). We use maximum-likelihood estimation to learn the optimal $\Gamma = \{\rho_{sr}, \theta_s\}$ from a training sample set. We assume the ρ_{sr} to be conditionally Gaussian distributed. The parameters of the Gaussian (mean and covariance matrices) for the ρ_{sr}, and the values for $\theta_s\}$ are estimated using an expectation-maximization technique.

In the expectation step, the posterior probabilities are computed using previous computed values for the $\Gamma = \{\rho_{sr}, \theta_s\}$. In the maximization step, the fusion parameters, $\Gamma = \{\rho_{sr}, \theta_s\}$ are recomputed based on the current posterior probabilities achieved in the E-step. The details of the algorithm are left out due to lack of space.

3.3. Fusion with Missing/Degraded Sensors

One important need of mobile sensing applications is the ability to handle missing sensory data, when a sensor dies due to a degradation of electronics, or runs out of power. Other techniques that employ a pool of features from multiple sensors cannot tackle missing or partial sensory information.

In our solution, we employ probabilistic mixture of experts, where each expert is dedicated to a single sensor. If a sensor is damaged and therefore does not give any data readout, we assign the posterior probabilities of that corresponding expert classifier for all classes to be equal, i.e. if sensor *s1* is degraded/damaged, the posterior probability for the experts $P_{s1,r}^{m}(x_{s1}) = 1/M$ for $m=1,\ldots M$, and $r=1,\ldots R$. Intuitively, this implies that the expert is unsure about the class of an observation if it's corresponding sensor is damaged. In this case, all the other sensory data inputs are used to make a final decision using Eq. (8) earlier.

4. MULTISENSOR ADAPTIVE CONTROL

A critical requirement in our distributed sensing system is the ability to detect and track important events while ensuring that certain system constraints are not compromised. This calls for an efficient control mechanism that takes into account all system resources, constraints, and measured/detected external events (derived from the sensor measurements) in order to control the system operation in real-time. The system controller needs to be adaptive in order to handle dynamic external events and changing conditions within the system. Adaptive control is of two types: (1) control in purely observation problems, and (2) control in systems with feedback.

From an application perspective, in health monitoring an observation problem corresponds to controlling sensor operational frequencies, making data storage, and transmission control decisions, while the person's medical condition changes dynamically. In greenhouse environmental monitoring, an observation problem could involve controlling sensor operational frequencies, making data storage, and transmission control decisions, while an external event occurs, such as a leak in a greenhouse. In feedback control problems, the control variables affect the external (measured and non-measured) variables and the external environment. In

health monitoring, this could correspond to injection of insulin to control a persons blood glucose levels. In greenhouse environment control, this corresponds to changing heat levels to ensure plant growth in the presence of a leak.

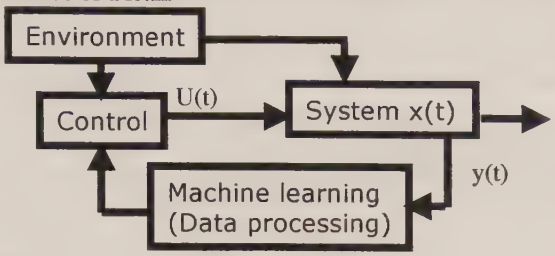


Figure 3: Control system representation of sensor network

Typically, real-world control optimization problems are characterized by constraints, multiple objectives, and dynamic properties. In particular real-world control problems are notoriously dynamic due to continuous changes in the external processes (environment) and, in many cases, the feedback between the controller and the controlled system. Additionally, changes in sensor sensitivities over time and in the internal system parameters makes the control problem more complex. A typical problem in hardware control is the drift of material parameters due to machinery wear-out and sensitivity to environmental changes, such as temperature, light, and humidity. Therefore, the applied control and optimisation technique should be able to continuously search for the best solution at frequent time intervals.

There are three main classes of control problems: (i) state stabilization (e.g. constant electricity supply of a power plant), (ii) system-to-system interaction (e.g. 'race' or 'war' with another system, such as pest control versus a crop pest), and (iii) profit maximization (e.g. market-oriented production optimization in a factory). **In this paper, we investigate control optimisation strategies in profit maximization.** Profit maximization involves measures that maximize a desired objective while minimizing the loss incurred in using system and environmental resources. The profit is a function of variables in the controller and in the controlled system.

We represent the interaction between the controller and the controlled system as shown in Figure 3. Here x(t) represents the internal state of the system at time t, u(t) is the control signal, and y(t) is the output from the system. The representation of a system requires that all variables that are either a part of the system or influence it indirectly be identified. Further, the type and range of each variable (control, system, or environment) needs to be determined. It is not always clear where to draw the line between control, system, and external environment; however, the following simple rules can be used to categorize variables into one of the three classes.

1. Control variables are variables that can be set by the controller.
2. System variables are directly affected by the control variables, but also through interactions with other (state and/or environmental) variables.
3. Environment variables represent components that influence the system, but are not directly affected by the control variables. They could be indirectly affected by changes in the control variables. *For purely observation problems, the control variables affect only the system, and not the environment.*

The change in system state is usually modelled by a number of difference equations of the form:

$$xi(t + h) = xi(t) + \Delta xi(u; x; v; t; h) \quad (3)$$

where xi is the ith system variable, $\Delta xi(u; x; v; t; h)$ is the update function, t is the time, h is the length of a time-step, and u, x, and v are the control signals, the system state, and the environment state of previous time-steps (sometimes several steps in the past). Many physical systems can be described by nonlinear differential equations, which can be approximated by difference equations using the Euler or the Runge-Kotta method [15]. In this case the update function $\Delta xi(u; x; v; t; h)$ of equation (1) is defined according to the used approximation method.

The exact form of the state update equations is dependent on the defined states for a problem, and their interrelationships. Our sensor network, however, has a problem-independent set of state (system) and control variables that we list below, For this paper, we consider the problem of determining the operational frequency of each sensor in the network and deciding on whether data should be stored/discarded (Table 1), given predefined system parameters (Table 3), current conditions of dynamic state variables (Table 2), and detected events (Table 4) by the pattern recognition system (Section 3). Together, they can be combined with the state/control/environment variables that are specific to the problem being considered to form a utility profit-loss function that should be optimised dynamically at each time step.

Control Variable	Description
Us	Frequency of data capture from sensor s
Uos	Operate/turn off sensor s
Ud	Discard/store sensor s data

Table 1: Sensor network control variables

State Variable	Description
PR	Remaining power
SS	Remaining storage

Table 2: Sensor network state variables

Fixed Variable	Description
PCs	Power needed for sensor s data capture
PXs	Power needed for sensor s data xmission
PPs	Power needed to run sensor s processing alg.
Ds	Sensor s data size (bytes)

Table 3: Sensor network predetermined variables

Classifier Dependent Variable	Description
IPCs	Information from sensor s (event/class label, labelling reliability, confidence)
Ecs	Class label criticality (benign, dangerous)

Table 4: Sensor network classifier-dependent variables

Utility	*Intuitive Description*	*Objective*	*Variable dependencies*
SIU	Sensor capture Info. Utility	Maximize	*FCs, IPCs, ECs*
SPC	Sensor capture power cost	Minimize	*FCs, PCs, PPs, PR*

Table 5: Profit/loss functions for control system

The controller that we use is a look-ahead predictive controller that takes into account the overall predicted profit of the controller over a future time horizon in choosing the control variables for the next time period. While several predictive control techniques exist, including model predictive control [16], fuzzy predictive control, and evolutionary algorithm-based control [12], we chose an EA-based technique due to the following reasons:

1. An explicit, correct model of the entire process is often not available, as in our case. MPC-based techniques typically require an accurate model of the entire process in order to function effectively.
2. Many iterative optimisation techniques only consider single solutions at a time. evolutionary algorithms maintain a population of solutions, rather than just a single solution. This provides the potential for a diversity of approaches to problem solving.
3. Evolutionary algorithms have particularly great potential to tackle dynamic problems compared to other iterative search techniques.

In the simplest case of an observation problem, we employ a single cost function C_R for the controller that maximizes a weighted positive sum of SIU, and weighted negative sum of SPC in the above table, with the constraint that Tr (PCs. FCs) < PR, where $Tr = T - Telapsed$ is the remaining minimum lifespan of the system. This ensures that the sensor network operates for a minimum total time of T. At each time instance, the optimal control settings that maximize the cost function are selected. Due to dynamic external events (in the

observed process), and changes within the system (decrease in power, storage, etc.), the control settings will adapt optimally for each situation.

For systems with external feedback, the cost function involves environment, state, and control variables, and its formulation is specific to the problem domain. For example, in greenhouse monitoring, the cost function is also a function of the profit in plant growth as a function of temperature, CO2, and humidity, while the loss is a function of the heat resource used to maintain a temperature range, water resource to maintain a humidity level, and CO2 resource used to ensure plant growth.

The optimisation algorithm used to determine the optimal control settings is a critical part of our control system. Real-time operability is essential for a deployable sensor network system. For optimal performance, we propose two phases of the GA-based control optimisation algorithm. In the design phase, we determined the optimal *M* operational conditions for various expected scenarios *offline*. In the operational phase, we first determine which of the *M* offline conditions correspond closest to the current (real-time) scenario, and use that as a starting point to initialise the GA controller. *This ensures swift convergence of the GA algorithm to the correct solution for optimal sensor operation and power management in the sensor network*. An analogue to our approach of using design control optimisation and real-time control optimisation already exists in the field of control theory. Real control problems are usually handled by either an offline design process or an online control strategy. Tuning a PID controller is a typical example of an offline problem. In online control, an adaptive control algorithm is repeatedly used to determine the best control signals during the control period.

The sensor network operational parameters must be updated at frequent intervals, especially when new events occur. The GA control optimisation routines should quickly (and with minimal iterations) determine the optimal control settings. GAs iteratively evaluate several possible solutions prior to convergence. The number of evaluations (#ev) can be balanced between population size (ps) and number of generations (gen), i.e., #ev = ps ¢ gen. For instance, 200 evaluations can be assigned as either ps = 200, gen = 1 or ps = 25, gen = 8. Our GA-based control optimisation algorithm is discussed briefly below.

```
Determine closest design operational setting
for (i=0; i < gen; i++ )
{       Crossover and mutation
        Evaluate each solution (cost) C_R for past, future H steps
        Selection
        Store best control operational policy
}
Select best control operational schedule for next Tc hours
```

5. SYNTHETIC DATA RESULTS

We consider synthetic data sets to show proof of concept of our mobile sensing algorithms. There are two main components in pattern recognition that are of concern in this paper, namely the dimensionality reduction component for reduced bandwidth, and the classifier component. In the real-time system control, we will show the results of our GA control optimisation on a greenhouse monitoring and control application with feedback. We shall illustrate results of each of these three components in our mobile sensing system. We consider simulated problems involving multiple sensors to illustrate our solution. This synthetic multisensory data corresponds to environmental data for multisensory environmental monitoring, or sensory data from individuals (physiological, pressure, etc.) for mobile health monitoring and quantitation applications.

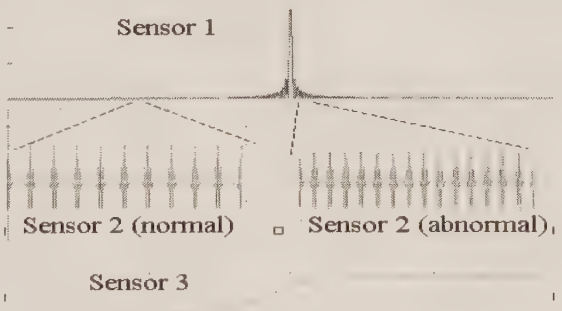


Figure 4: Ideal sensor data with no noise over a time period

For the health monitoring scenario, we consider three (synthetic) sensors that are derived from models of physical sensors for real-time detection of anomalies/events in an individual's health. These sensors provide information about the physiological condition of a person, and also metabolic changes in the person via interstitial-fluid (ISF) measurements [13]. These could correspond to real-time glucose measurements (which could be used for diabetic monitoring towards the goal of insulin control), estimating endurance strength of soldiers in the battlefield, or ethanol estimation which along with other physiological measures could be used for mobile alcohol monitoring. This has potential uses for monitoring of workers in critical jobs (maritime professionals, nuclear power plant operators) or to ensure that drivers with prior drunken violations remain sober at all times. In these prior health monitoring examples, the requirement is to monitor health for extended periods of time (weeks) in extremely low bandwidth, thus making automated decision making and resource management a critical issue. While we use the mobile health monitoring example as a case-study, our method is easily applicable to other scenarios such as environmental monitoring, plant growth monitoring, etc. We use models of physical sensors to generate realistic synthetic data. The performance of the system under varying levels of sensory noise (i.e. increasing sensor degradation) is evaluated. The two synthetic physiological measurement sensors are heartbeat (sensor 2), and differential blood pressure readouts (sensor 1). An ethanol

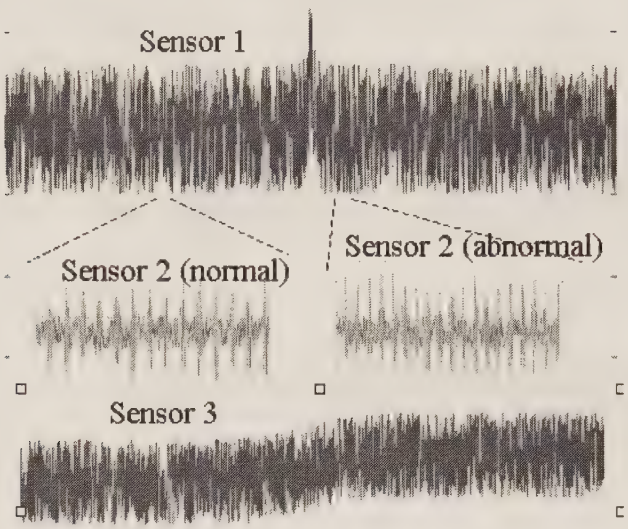


Figure 5: Sensor data with high noise over a time period

detector that has a slow response curve to ingestion of alcohol is also used (sensor 3). The noise-free signals for each sensor is shown in Figure 4.

For the greenhouse monitoring and control application the simulated sensor data could correspond to temperature, humidity, and/or CO2 levels.

Fig 6: (a) Sensor 2 readout (normal), and (b) abnormal event at high-noise levels (+/- 0.8)

The sensor data in Figure 4 is noise-free. In reality, the sensor readouts are corrupted by noise, due to the data capture process, or due to the interface electronics (analog to digital conversions, quantization, SNR of the physical process). We model the noise in the following manner. For sensors 1 and 3, where the abnormal events are characterized by a peak (due to the differential nature of the readouts) and a ramp signal, we consider additive noise. The additive noise is uniformly distributed random noise. For sensor 1, the noise varies from U[-0.1,0.1] to U[-0.8,0.8]. For sensor 2, the additive uniform noise varies from U[-0.1,0.1] to U[-7,7]. For the heartbeat sensor (Sensor 2), the noise is both additive in amplitude and additive in frequency. The amplitude noise for sensor 2 varies from U[-0.1,0.1] to U[-0.8,0.8]. The frequency noise for sensor 2 varies

from U[-0.1e+3,0.1e+3] to U[-0.8e+3,0.8e+3. Representative signals for high noise levels are shown in Figure 5.

5.1. Dimensionality Reduction Results

The MRDF dimensionality reduction scheme has been proven to yield highly efficient reduction/compression of 2-D image input data in prior work [4,5] for both, data classification and data class representation. In our current application, we are interested in classifying data initially. Features that represent the sensor data well (such as PCA) could be useful in low-bandwidth scenarios where the user needs to see much/all of the raw sensory data.

We need to evaluate the performance of the MRDF feature extraction procedure under varying levels of noise. We use two measures in evaluating the performance of our MRDF dimensionality scheme: we measure the compression of the raw sensor data, and the ratio of the inter-class separation of the noisy data in feature space versus the inter-class separation of the noise-free signals.

The compression efficiency of the MRDF is 2250% since raw (8-bit integer) sensor data over a time period of 360 sec is compressed to 4 floating point (32 bit) MRDF features. This is a reduction by a factor of 22.5.

The usefulness of the MRDF in detecting important events is measured by the following measure: $E_{DR} = E_{D\_Noisy}/E_{D\_Ideal}$ for different noise levels. This ratio is given in Table 6. The inter-class separation is excellent for low-to-medium noise, and degrades gracefully at higher noise levels. At the higher noise levels, the inter-class separation falls largely due to the error caused by the frequency change for normal vs. abnormal cases (see Fig. 6); in this case, a normal event (Fig. 6a) is easily mistaken for an abnormal occurrence (Fig. 6b) due to the noise in the oscillation frequencies and signal amplitude.

NOISE LEVELS (+/- Range, Uniform RV)				E_{DR} (Avg).
Sensor 1	Sensor 2 (Amplitude)	Sensor 2 (Frequency)	Sensor 3	
0.1	0.1	0.1e+3	0.1	0.97
0.5	0.5	0.5e+3	0.5	0.82
0.8	0.8	0.8e+3	0.8	0.4

Table 6: Inter-class separability of dimensionality reduction

5.2. Multisensor Classification and Fusion Results

The multisensor fusion and classification results at various levels of noise are presented next. To ensure statistically significant classification results, we used a test set of 1000 samples. The classification results for different noise levels are shown in Table 7. As expected, the classification performance reduces with increasing noise. The classification performance of 86.1% at the highest noise levels is good, given the deterioration in the SNR, as evident from Fig. 6.

NOISE LEVELS (+/- Range, Uniform RV)				P_C (TOTAL)
Sensor 1	Sensor 2 (Ampl.)	Sensor 2 (ω)	Sensor 3	
0.1	0.1	0.1e+3	0.1	100%
0.5	0.5	0.5e+3	0.5	96.5%
0.8	0.8	0.8e+3	0.8	86.1%

Table 7: Pc for varying noise levels with all 3 sensors functional

5.3. Multisensor classification with missing sensors

We now discuss the case where one sensor degrades/stops working, and show that our classification scheme handles such a deterioration very gracefully. From our tests, we found Sensor 3 (synthetic ethanol sensor) to be the most robust to noise. We evaluated the performance when Sensor 3 stops, and only data from Sensors 1 and 2 are available. The posterior probabilities for each class for Sensor 3 are set to 0.5 (since we have only two classes). The classification results for this case are shown in Table 8, again at varying noise levels as before. In this case, the performance was good for low-medium noise levels. At the highest noise level, the performance decreased significantly due to the fact that sensor 2 often has a significantly incorrect frequency for both classes at this noise (see Fig. 6).

NOISE LEVELS (+/- Range, Uniform RV)				P_C (TOTAL)
Sensor 1	Sensor 2 (Ampl.)	Sensor 2 (ω)	Sensor 3	
0.1	0.1	0.1e+3	x	100%
0.5	0.5	0.5e+3	x	93.5%
0.8	0.8	0.8e+3	x	79.7%

Table 8: Pc for varying noise levels with one dysfunctional sensor.

5.4. Multisensor Control Optimisation Results

We tested our control optimisation algorithm on a problem with feedback. Specifically, we considered the greenhouse monitoring and control case, where along with system (storage and power) control, we also control plant growth in a greenhouse while ensuring that system resources such as power, heating, water, and CO2 are not over utilized beyond the system capacity. The relation between the different variables (control, system and environmental/plant variables) in greenhouse monitoring and control is shown in Fig. 7.

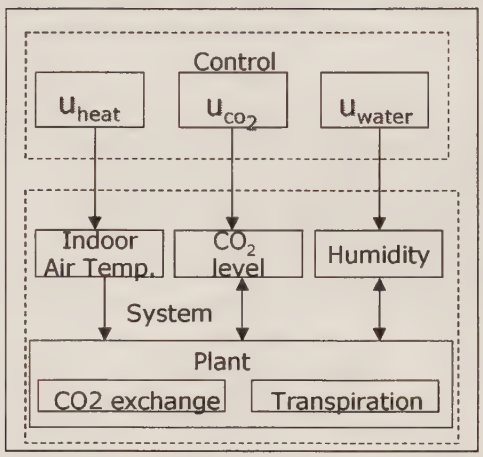


Figure 7: Greenhouse model for control optimisation

To test our algorithms, we used an opensource greenhouse simulator called Garden with Insight (GWI) based on the well-known erosion-productivity impact calculator (EPIC) – an agricultural management tool, and the code provided in [14]. We tested the results of our GA control algorithm on plant growth in the simulator over a six week time period. The biomass growth during this period with GA-based control optimisation is shown in Fig. 7. The plant growth achieved when the GWI greenhouse plant simulator was run without any adaptive control was observed to yield significantly lower plant productivity, as seen in Fig. 8. In fact, adaptive

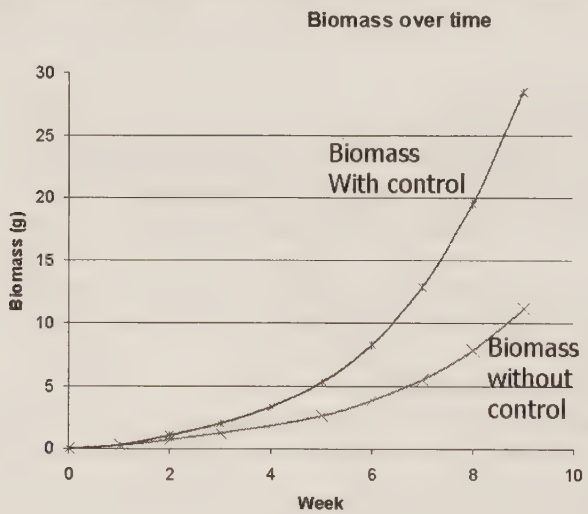


Figure 8: Plant growth over a 6 week period with GA-based control optimisation and without any control.

control of the greenhouse yielded a 2.5x increase in productivity (250% better performance) than running the greenhouse without any feedback control. These initial results hint at the immense potential of our control optimisation in optimally controlling a multivariable process under dynamically changing conditions for a variety of applications.

6. CONCLUSIONS AND FUTURE WORK

We have proposed and shown the utility of a new dimensionality reduction technique and a classification approach from features extracted from multiple sensors while handling missing sensory data robustly. Evolutionary Genetic Algorithms are used to determine the best operational/control schedule. Our current tests shows that our approach has extremely good potential on simulated greenhouse data. Further tests on real data under a variety of situations remain to be done. Extensions of the classification algorithms to allow data prioritisation for passing critical information in very-low bandwidths is an important next step. Techniques to adapt the classifiers under slowly changing environments is a necessity. While a number of challenges remain prior to realizing our goals, we believe that our solutions for machine learning and adaptive controls are a critical step towards this endeavour.

7. ACKNOWLEDGMENTS

The research described in this paper was carried out by the Children's Hospital Los Angeles/Univ. of Southern California. This work has been sponsored in whole with Federal Funds from the National Institute on Alcohol Abuse and Alcoholism, National Institute of Health, Department of Health and Human Services under Contract No. N01AA33004.

8. REFERENCES

[1] intel.com/research/exploratory/heterogeneous.htm
[2] System architecture directions for network sensors, Jason Hill, Robert Szewczyk, et. al. ASPLOS 2000, Cambridge, November 2000 [Bellutta00]
[3] S.S. Wilkes,Mathematical Statistics, Wiley, New York,1962
[4] A. Talukder and D. Casasent, "A closed-form neural network for discriminatory feature extraction from high-dimensional data ", Neural Networks , Vol. 14 (9), Pages 1201-1218, November 2001.
[5] D. Casasent and A. Talukder, "Nonlinear features for robotics, inspection, and face recognition", Keynote address, Proc SPIE, July 1999.
[6] Frank Dellaert, The Expectation Maximization Algorithm, College of Computing, Georgia Institute of Technology, Tech. Report GIT-GVU-02-20, Feb. 2002
[7] N. Vlassis and A. Likas. A greedy EM algorithm for Gaussian mixture learning. Neural Processing Letters, 15(1):77-87, February 2002.
[8] Jordan, M. I., & Jacobs, R. A. Hierarchical mixtures of experts and the EM algorithm. Neural Computation, 6, 181-214, 1994
[9] Gagliano, R. A., M. D. Fraser, and M. E. Schaefer: 1995, `Auction allocation for computing resources'. Communications of the ACM 38(6), 88--102.
[10] Lu, J. A. Stankovic, G. Tao, and S. H. Son. Feedback control real-time scheduling: Framework, modeling and algorithms. Journal of Real Time Systems, 2001. Special Issue on Control-Theoretical Approaches to Real-Time Computing
[11] Zhao, et. al., Preemptive scheduling under time and resource constraints
[12] A. J. Chipperfield and P. J. Fleming. Multiobjective gas turbine engine controller design using genetic algorithms. IEEE Transactions on Industrial Electronics, 43(5), October 1996.
[13] Eppstein, et. al., "Microporation of human skin for monitoring the concentration of an analyte", US Patent US5885211, March 23, 1999.
[14] Rasmus K. Ursem, Thiemo Krink, Bogdan Filipic, "A Numerical Simulator of a Crop-Producing Greenhouse", EVALife Technical Report No. 2002-01, 2002.
[15] Robert Adams, "Calculus – A complete course", Addison-Wesley publishers, 3rd edition, 1995.
[16] Eduardo Camponogara, Dong Jia, Bruce H. Krogh, and Sarosh Talukdar, "Distributed Model Predictive Control", IEEE Control Systems Magazine, pp. 44-52, February 2002

Characterization of photorefractive deconvolution techniques for one-way image transmission through an aberrating medium

B. Haji-Saeed, S. K. Sengupta, M. Testorf, D. Megherbi and W. Goodhue
Physics and Electrical and Computer Engineering Departments,
University of Massachusetts Lowell, Lowell MA 01854

J. Khoury and C. L. Woods
Air Force Research Laboratory / SNHC, Hanscom Air Force Base, MA 01731

J. Kierstead
Solid State Scientific Corporation, Hollis NH 03049

ABSTRACT

In a previous publication we introduced a new photorefractive four-wave mixing deconvolution, FWMD, image correction approach for achieving one-way image transmission through an aberrating medium. In this paper we extend our work to include additional image degradations and more test cases. We characterize the performance as a function of the input beam ratios for four metrics: signal-to-noise ratio (SNR), normalized mean-square-error (NMSE), edge restoration (ER), and the peak-to-total energy ratio (PTE). In our characterization we color-code the best beam-intensity ratio 2D region(s) for each of the above metrics. Test cases are simulated at the optimal values of the beam-intensity ratios.

Keywords: Image processing, optical processing, Distortion correction, image restoration, adaptive optics, photorefractivity, four wave mixing.

1. INTRODUCTION

The correction of image distortions produced by transmission through various aberrating media constitutes a very important class of problems in diverse fields of human endeavor. Practical sources of such image corruption include atmospheric turbulence in both astronomical and satellite imaging [1,2], and scattering by biological tissues in medical imaging. Researchers have expended considerable effort to find a convenient way to restore the original unblurred image from these distortions and to reduce any associated noise as much as possible.

An image transmitted through the atmosphere suffers distortion from two different mechanisms: turbulence and scattering. Atmospheric turbulence can be described as a random phase aberration. It means that these aberrations change in time frequently and also change the point spread function (PSF), which is needed to characterize the image [1,2]. The atmospherically distorted image can be visualized as a crumpled sheet of paper, which is smooth and flat before entering the atmosphere. The wave front that enters the atmosphere comes apart. Some parts of it go through hotter than average air and other parts are retarded and the plane wave front is deformed.

In a previous publication [3], we have proposed and demonstrated a new technique that provides improved image correction. Our new approach uses real-time holography to produce four-wave mixing in a photorefractive crystal for adaptive one-way image transmission through an aberrating medium. We also compared this real-time holographic method with other wave-front correction approaches like phase or amplitude coding using spatial light modulators [4,5], double-pass correction using phase conjugate mirrors [6], and direct correction using deformable mirrors (DM) like those in telescopes [1,2]. In our previous experimental demonstration, we used both BGO (Bismuth Germanium Oxide) and $BaTiO_3$ (barium titanate) photorefractive crystals. These crystals provide good resolution, and they are inherently massively parallel. Photorefractive materials provide moderate to fast response that can be adjusted by changing the light

Optical Pattern Recognition XV, edited by David P. Casasent, Tien-Hsin Chao, Proc. of SPIE Vol. 5437 (SPIE, Bellingham, WA, 2004) · 0277-786X/04/$15 · doi: 10.1117/12.548076

intensity, laser wavelength, photorefractive material, applied electrical field and other experimental parameters. Faster variations in the image can be accommodated at important near-infrared atmospheric transmission windows by using semiconductor quantum well photorefractive materials, including those based on quantum dots. For example, Fe-doped InGaAs/InGaAsP has been successfully used at 1.55 μm to provide microsecond response times, which provides very fast 2D signal or image processing speeds [7].

In this work we extend our photorefractive four-wave mixing deconvolution approach [3] by performing an extensive computer simulation to find the optimal beam-intensity ratio operating point and region for each of the following metrics: SNR, NMSE, ER and PTE. The figures of the optimal restored image for each of these metrics shows the performance at the optimal operating values of the beam-ratios for that metric. The color-coded figures of the 2D surface for each performance metric show the best to the poorest operating regions in terms of two beam-intensity ratios.

2. BACKGROUND

When two write beams A_1 and A_3 corresponding to the signal and reference beams, respectively, mix with one read-out beam A_2 counter-propagating to A_1 within a photorefractive crystal, a phase-conjugate beam A_4 is generated via degenerate four wave mixing (DFWM). This basic scheme, shown in Figure 1, provides the basis for all forms of image correction by photorefractive deconvolution via four-wave mixing, including phase correction, adaptive amplitude correction and noise filtering. We provide this combination of image restoration by using a design in which, the signal beam A_1 contains the information of the aberrator impulse response, which is Fourier transformed on the crystal, and read-out beam A_2 is a plane-wave. The reference beam A_3 is the aberrated beam and contains has information of the convolution between the image and the aberrator impulse response. The phase-conjugate beam A_4 is the corrected beam, whose beam amplitude, as it exits the photorefractive crystal, is given by the formula [3]:

$$A_4(\nu_x,\nu_y)=\frac{\left((A_1/\lambda f)H(\nu_x,\nu_y)(A_3^*/\lambda f)H^*(\nu_x,\nu_y)F^*(\nu_x,\nu_y)\right)A_2/\lambda f}{I_1|H(\nu_x,\nu_y)|^2/(\lambda f)^2+I_3|H(\nu_x,\nu_y)|^2|F(\nu_x,\nu_y)|^2/(\lambda f)^2+I_2/(\lambda f)^2} \quad (1)$$

Here A_1, A_2, A_3 are the amplitudes; I_1, I_2, I_3 are the beam intensities; H (ν_x, ν_y) is the Fourier transform of h (x, y), the aberrator impulse response function; F (ν_x, ν_y) is the Fourier transform of f (x, y), the object (input image) function; λ is the laser wavelength; and f is the focal length of the lenses L.

If we define the beam-intensity ratios, $m=I_1/I_3=|A_1|^2/|A_3|^2$, and $p=I_2/I_1=|A_2|^2/|A_1|^2$, then Eq. (1) becomes:

$$A_4=\frac{\sqrt{m}\,F^*(\nu_x,\nu_y)H^*(\nu_x,\nu_y)H(\nu_x,\nu_y)}{m|H(\nu_x,\nu_y)|^2+mp+|F(\nu_x,\nu_y)|^2|H(\nu_x,\nu_y)|^2}(A_2/\lambda f) \quad (2)$$

This form will be utilized in calculating image correction parameters in terms of the beam-intensity ratios.

This adaptive aberration correction has the following limiting forms:

a) For a white frequency pure phase aberration where $|H|^2 \sim \alpha$ with a small m and a large mp the approximate corrected beam becomes:

$$A_4(\nu_x,\nu_y)=\alpha\frac{\sqrt{m}\,F^*(\nu_x,\nu_y)}{mp}(A_2/\lambda f) \quad (3)$$

b) For complex aberrations caused by motion or misfocusing and when m is large and mp is very small, the approximate corrected beam becomes:

$$A_4(\nu_x,\nu_y)=\frac{\sqrt{m}\,F^*(\nu_x,\nu_y)}{m}(A_2/\lambda f) \quad (4)$$

Additional information about the complete derivation of these equations may be found in reference [3].

3.COMPUTER SIMULATIONS

We developed computer simulations for four different inputs: 1) a binary three-bar pattern of size 24x120 pixels that has both step function and square function characteristics, 2) a binary two dimensional comb function of size 128x128 pixels in which a one-pixel line of unit pixels is separated by a line of four zero pixels, 3) a delta function using a unit pixel of size 1x1, and 4) a standard gray-scale image of a human face of size 120x120. In this paper we only discuss the three-bar pattern and the point source cases. The results of the other two inputs will be presented elsewhere.

The different aberration functions used for distorting these inputs were: 1) a phase-only aberrator (atmospheric aberration), 2) a circular aberrator (misfocusing), 3) a narrow rectangular aberrator (motion aberration), and 4) a combination of the last three distortions. As an alternative for atmospheric aberration we used a uniformly distributed random phase-only function of size 120x120, the misfocusing impulse response is a circle with a radius of 5 pixels of ones, the motion impulse response is a rectangle of 2x50 pixels array of ones. The aberrators and input are located in the middle of our total array, which is a 256x256 zeros (null) array. The inputs and the aberration functions are shown in Figures 2 and 3, respectively.

Convolving the input with one of these aberrators plus noise creates our blurred images. The noise is created by a random number generator, which produces Gaussian noise with zero mean and variance of one. Different noise levels have been produced by multiplying by σ =100, 25, 10, 1 and 0.1. The aberrator function and signal plus noise, which include the blurred image and noise, are Fourier transformed. The signal plus noise is conjugated and inserted in Eq. 2 as $H^*(v_x, v_y)$ $F^*(v_x, v_y)$ and the Fourier transform of the aberrator function is brought in as $H(v_x, v_y)$.

We select the SNR, NMSE, ER and PTE as a set of quantitative metrics for evaluating one of our images corrections. In our calculation, SNR is defined as the ratio of the mean (average) to the variance in intensity, NMSE is the rms of the difference between the normalized output and normalized input. We define the edge restoration criterion, ER, as the ratio of the edge-energy of the corrected (output) image to the edge-energy of the blurred (input) image. Using the middle bar of the three bar pattern, the blurred bar area is the extended by the point spread function of the circle, and this is the computational area, CA, for calculating the ER. This edge-energy is defined as the energy in the computational area less the energy within the region of the perfect bar area. Our definition of ER is shown in Eq. 5, where both the restored edge and blurred edge energies are normalized by their respective energy in the CA.

$$ER = \frac{\sum |\text{Blurred\_image\_edge}|^2 \Big/ \sum |\text{Blurred\_image}|^2}{\sum |\text{Restored\_image\_edge}|^2 \Big/ \sum |\text{Restored\_image}|^2} \qquad (5)$$

And we define our last quantitative metric PTE (used as an alternative to the Strehl ratio since photorefractive optical wave mixing is a nonlinear system) evaluated for our point source input by

$$PTE = \frac{|\max(\text{FFT\_output})|^2}{\sum |\text{FFT\_output}|^2} \qquad (6)$$

Figure 4 shows the results of our computer simulation for the three-bar pattern input. The first, second and third columns show the results for rectangular aberrator (motion aberration), phase-only aberrator (atmospheric aberration) and circular aberrator (misfocusing) respectively. Figure 4(A) is the blurred three-bar pattern distorted by motion aberration plus noise. Figure 4(B) is the corrected image after photorefractive filtering at beam ratios $m=10^5$ and $p=10^6$ and Figure 4(C) is the difference between the original and corrected images.

Figure 4(A') is the blurred image with a phase-only aberrator and additive noise. In this figure the mean (average) of the information over the entire array is approximately 75 and Figure 4(B') shows the corrected image after photorefractive filtering at beam ratios $m=10^7$ and $p=10^7$. Operating conditions at these beam ratios correspond to the undepleted pump approximation where it is possible to produce an amplified phase conjugation beam. Figure 4(C') is again the difference between these two images.

Figure 4(A'') demonstrates the blurred image with additive noise caused by convolution with a circ function to present a case of misfocusing. The SNR for this image before correction is 4.4. Figure 4(B'') is simulating the corrected image with photorefractive filtering at beam ratios $m=10^5$ and $p=10^8$. Under these conditions, the reflectivity of the phase conjugator is extremely small, and may require an image intensifier to observe the images. Figure 4(C'') is the difference

between the last two images. After image restoration, the measurements for our three metrics were evaluated using the middle bar of the three-bar array. The SNR, NMSE and ER are equal to 3.9, 7108 and 2.38, respectively. The image was completely restored at the expense of a slight degradation in signal to noise ratio. Therefore it is clear that our technique can correct the image from these extreme aberrations. Our point source input results are demonstrated in Figure 5. The first row shows the blurred point source distorted by motion aberration, phase only aberrator and misfocusing, respectively. The second row demonstrates corrected images from these aberrations and the third row is the differences of the original image and the corrected one.

We performed an extensive computer simulation to find the optimal operational point in terms of SNR, NMSE, ER and PCE. These measurements were taken only for the circular aberration. The color-coded 2D surfaces in Figures 6 and 7 for each performance metric show the best to the poorest operating regions in terms of beam-intensity ratios. of 10^{-10} to 10^{10} for m (x-axis) and 10^{10} to 10^{-10} for p (y-axis) at each noise level. The results are presented in the form of false color graphs. It is clear from Figure 6 that our results are at the optimal beam ratio operating point. The first, second and third rows in Figure 6 show the performance of SNR, NMSE and ER for different levels of noise. Noise levels are defined by different rms values of σ where the first, second and third columns correspond to σ=1,10 and 25, respectively. The darkest area has the maximum value in terms of our quantitative metrics and the x- and y-axes are the beam ratios m and p, respectively. This picture clearly demonstrates that the optimal beam ratios for all of our metrics are almost the same. Therefore, if we work in this regime we have the most SNR, the least NMSE and the most ER.

Our other quantitative metric is the peak to total energy (PTE). The performance of this criterion is shown in Figure 7 for noise levels of σ =1, 10 and 25.

4.CONCLUSIONS

Our previous publication introduced a new photorefractive four-wave mixing deconvolution approach for one-way image transmission through an aberrating medium. This technique has the ability to simultaneously correct both amplitude and phase distortions. Our procedure uses the interference-pattern convolution between the blurred image and the probe beam wave fronts to remove the distortion induced by the aberrating medium. Extremely fast real-time image correction can be carried out at near-infrared wavelengths by operating at atmospheric transmission windows and using semiconductor quantum well photorefractives extend our earlier work to include more test cases. For this work we carried out extensive computer simulations on new test cases and included additional image degradations to demonstrate our applications. We characterize the image restoration performance as a function of the input-beam-intensity ratios for four metrics: signal-to-noise ratio, SNR;, normalized mean square error, NMSE; edge restoration, ER; and the peak-to-total-energy ratio, PTE. In this characterization, the image restoration test cases are simulated at the optimal values of the beam-intensity ratios, and the best beam-intensity ratio 2D region(s) for each of the above metrics is color-coded. The false color multidimensional data visualization technique identifies the conditions that will give us the best performance in terms of 1) large values of SNR, ER and PTE and 2) the least error in terms of NMSE.

REFERENCES

[1] Francois Roddier, "Adaptive optics in astronomy", Cambridge University Press (1999).
[2] V. P. Pauca, B. L. Elerbroek, R. J. Plemmons, X. Sun, "Structure matrix representations of two parameter.Hankle transforms in adaptive optics," Linear Algebra and its applications 316, 29-43, (2000).
[3] B. Haji-Saeed, D. Pyburn, R. Leon, S. K. Sengupta, W. Goodhue, M. Testorf, J. Kierstead, J. Khoury and C. L. Woods, "Photorefractive Nonlinear Deconvolution For One-Way Image Transmission Through Aberrating Media", Optical Pattern Recognition XIV, Proceeding of SPIE Vol.5106, p186 (2003).
[4] J. Khoury, J. Fu and C. L. Woods, "Phase-coding technique for one-way image transmission through aberrating media", Opt. Lett. 19,1645 (1994).
[5] J. Khoury, J. Fu, C. L. Woods, "Phase coding technique for signal recovery from distortion," Optical Eng., 34, 840 (1995).
[6] D. M. Pepper, Amnon Yariv "Compensation for phase distortions in nonlinear media by phase conjugation," Opt. Letts, 5, 59 (1980).
[7] C. De Matos, A. Le Corre, H. L Haridon, S. Gosselin and B. Lambert," Fe-doped InGaAs/InGaAsP Photorefractive multiple quantum well devices operating at 1.55 μm," Appl. Phys. Lett. 70 (26), 3591(1997).

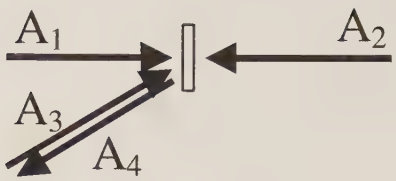


Fig. 1. Four-wave mixing in a photorefractive medium

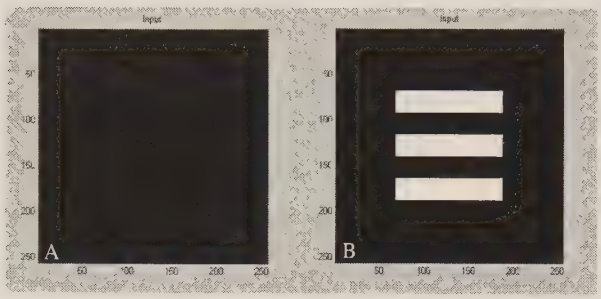


Fig. 2. Binary test inputs: A) Point source (single pixel in center), B) Three-bar pattern

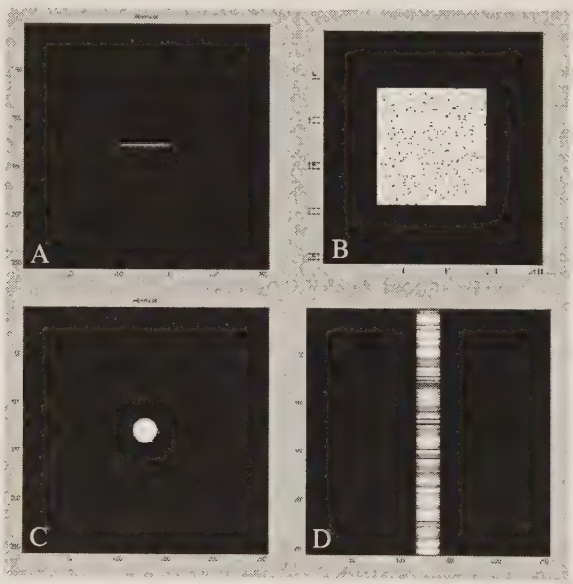


Fig. 3. Aberrators: A) motion (50x2 rectangular binary blur); B) phase-only (128x128 uniformly distributed random phase aberrator); C) misfocusing (circular blur with a radius of 5); and D) combination of the last three aberrators

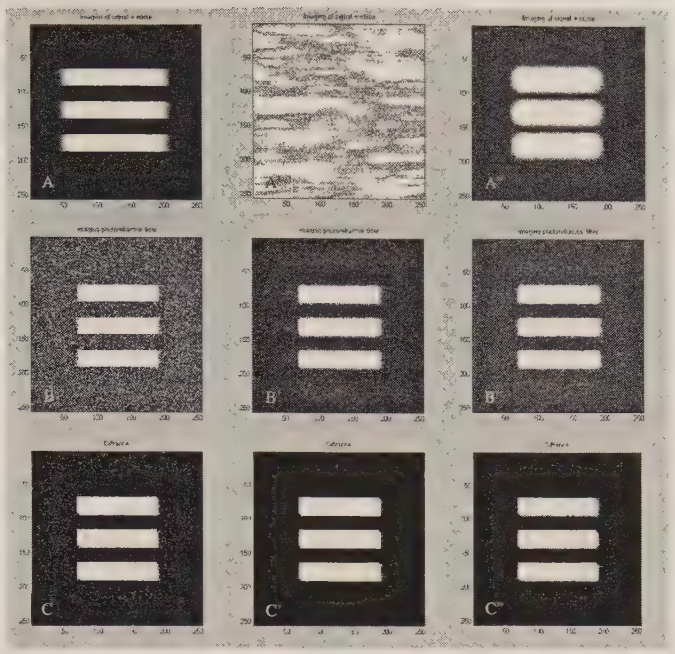


Fig. 4. Bar image restoration tests: A) three-bar pattern with motion blur plus noise, B) corrected image, C) the difference between the original and corrected images; A') blurred three-bar pattern with phase only aberrator plus noise; B') corrected image; C') the difference between the original and corrected images; A") blurred three-bar pattern with circular blur plus noise; B") corrected image; C") the difference between the original and corrected images.

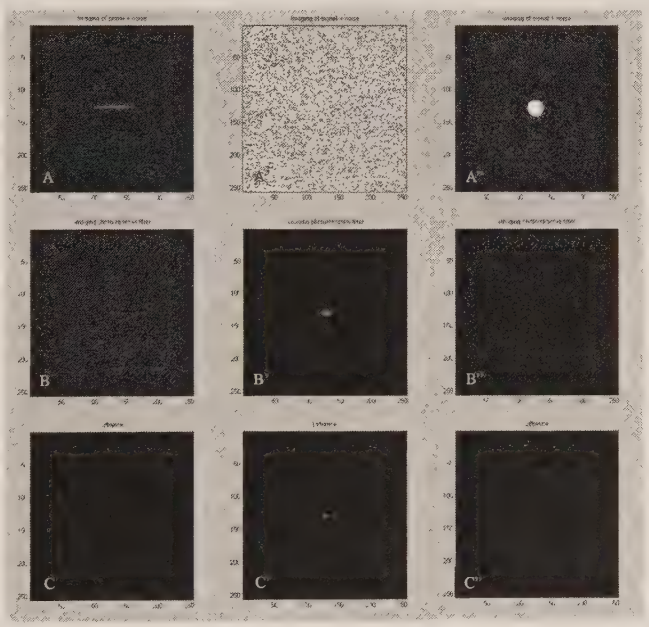


Fig. 5. Point source image restoration tests: A) blurred point source with rectangular aberrator plus noise, B) corrected image, C) the difference between the original and corrected images; A', Blurred point source with phase only aberrator plus noise, B') Corrected image; C') The difference between the original and corrected images; A") Blurred point source with circular blur plus noise; B") Corrected image, C") The difference between the original and corrected images.

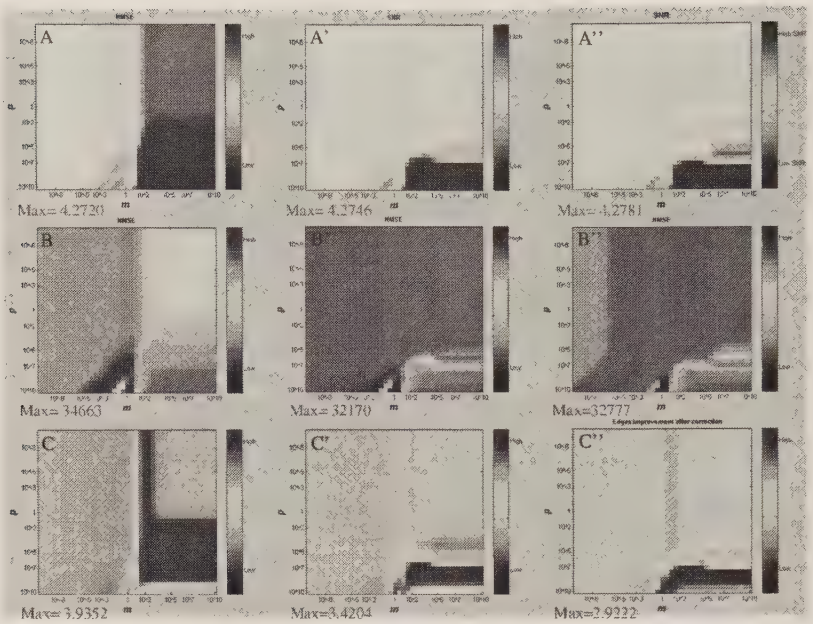


Fig. 6. Color-coded metric performance plotted versus beam intensity ratios m and p for the x-axis (10^{-10} to 10^{10}) and y-axis (10^{10} to 10^{-10}), respectively (increasing I_1 only produces a line with a slope of 1): The first row is the performance of SNR for different noise levels A) σ=1, A') σ=10, A'') σ=25; The second row is NMSE for B) σ=1, B') σ=10, B'') σ=25 and the third row is the edge restoration for C) σ=1, C') σ=10, C'') σ=25.

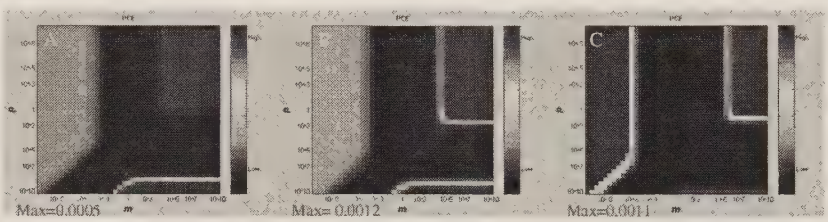


Fig. 7. The performance of PCE for different noise levels: A) σ=1, B) σ=10, C) σ=25.

Resolution Limits in Imaging LADAR Systems

J. Khoury, C. L. Woods, J. Lorenzo
Air Force Research Laboratory / SNHC, Hanscom Air Force Base, MA 01731

J. Kierstead
Solid State Scientific Corporation, Hollis, NH 03049

D. Pyburn and S. K. Sengupta
University of Massachusetts, Lowell, MA 01854

ABSTRACT

In this paper, we introduce a new design concept of laser radar systems that combines both phase comparison and time-of-flight methods. We show from signal to noise ratio considerations that there is a fundamental limit to the overall resolution in 3-D imaging range laser radar (LADAR). We introduce a new metric, volume of resolution (VOR), and we show from quantum noise considerations, that there is a maximum resolution volume, that can be achieved, for a given set of system parameters. Consequently, there is a direct tradeoff between range resolution and spatial resolution. Thus in a LADAR system, range resolution may be maximized at the expense of spatial image resolution and vice versa. We introduce resolution efficiency, η_r, as a new figure of merit for LADAR, that describes system resolution under the constraints of a specific design, compared to its optimal resolution performance derived from quantum noise considerations. We analyze how the resolution efficiency could be utilized to improve the resolution performance of a LADAR system. Our analysis could be extended to all LADAR systems, regardless of whether they are flash imaging or scanning laser systems.

Optical Pattern Recognition XV, edited by David P. Casasent, Tien-Hsin Chao, Proc. of SPIE Vol. 5437 (SPIE, Bellingham, WA, 2004) · 0277-786X/04/$15 · doi: 10.1117/12.548077

1. INTRODUCTION

LADAR (Laser Radar) systems, particularly imaging LADAR systems are seeing increased use in a variety of applications from aerial mapping to robotic vision. In these varied applications, one of the most important operating parameters, as in any imaging system, is the resolution of the system. One of the principal reasons for utilizing lightwave frequencies is the increased resolution that is available as compared to microwave radar. Currently, there is a lot of work going on to improve the quality of images obtained from 3-D imaging LADAR systems. Our analysis indicates that simply increasing the number of sampling points in a scanning laser system or building an imaging array with a very large number of pixels to increase the image resolution, may be a waste of effort, since this comes only at the expense of range resolution.

In this paper we introduce a new operational principle of LADAR that combines simultaneously phase comparison and time-of-flight. Phase comparison provides high resolution[1] and cw (continuous wave) or quasi-cw LADARs utilizing it are suitable for use with low power diode laser sources[2, 3]. We show from quantum noise considerations that there is a fundamental limit on the total resolution that can be achieved in 3-D imaging LADAR. We introduce a new metric, volume of resolution (VOR) that describes the resolution of 3-D images. We derive accurate formulae for VOR under both quantum noise limits and other system constraints, e.g. geometrical optics considerations accompanied by A/D conversion.

2. THEORY

In this section, firstly we will summarize the operational principle of a conventional phase comparison LADAR, then we will introduce a new low frequency LADAR in which its operational mechanism combines both heterodyne detection and time of flight.

2.1 Conventional Phase Comparison Laser Radar

Figure 1 (A) shows a block diagram of a laser radar using phase comparison. A laser source is amplitude modulated at a modulation frequency ω. A target is scanned by the laser beam through an X-Y scanner and suitable transmitter optics. The received return signal is collected by appropriate receiver optics and amplified in an optical amplifier. The amplified optical signal is input to a photodetector, whereby the modulation signal is recovered and then electrically amplified. We are using incoherent or 'direct' detection. The amplified detector signal is now homodyne detected in a mixer/phase detector, by mixing with the same modulation frequency and low-pass filtered to provide the range information. This signal S is now digitized in an A/D converter for further processing and display. The basic operating principle is thus the same as that for cw microwave radars[4].

If P_0 is the input optical power into the LADAR receiver, after phase comparison, the LADAR range return voltage signal S, is given by:

$$S = 2E(GP_0 e\eta / h\nu) R_l \sin(2\omega R / c) \qquad (1)$$

where G is the gain of the optical amplifier in the return receiver, e is the electronic charge, h is Planck's constant, η is the quantum efficiency of the detector, ν is the frequency of the LADAR laser, ω is the modulation frequency, R is the range, and c is the velocity of light. R_l is the detector load resistance. The eη/hν term is the responsivity of the detector.

2.2 Low Frequency Gated Laser Radar System

The operational mechanism of this laser radar combines both phase comparison and time-of-flight range measurement. This scheme, shown in Fig. 1(B), essentially is similar to the previous one, except that a second mixer is introduced, in which one of its inputs is the output of the first mixer and its second input is a square waveform coming from the function generator. The range information is extracted from digitizing the phase comparison output in the linear range of the operational curve. The range information extracted from digitizing the phase comparison output in the linear range of operation, corresponds to the range interval between R1 and R2. This interval, shown in Fig. 2, corresponds to a distance along the curve of Pi/4 to –Pi/4, giving us a 90% accuracy in this linear range. The gating of the phase comparison signal from the first mixer with the square pulse from the square function generator gives the possibility of extracting the required range interval.

To satisfy this condition the square pulse generator should be active for time of t of $Pi/4\omega$, while the full period of the square modulation should be Pi/ω, which is basically equivalent to the duration time on each pixel through the scanning process.
For scanning over all of the linear range, it is necessary to have a square function with a 50% duty cycle. The extent of the volume to be scanned is determined by the scanner capability to scan in the X and Y directions, and the extent of the range to be scanned is determined by the laser modulation frequency. On the other hand, the scanned range is determined by the duty cycle of the square function generator. The smaller the duty cycle, the smaller the range scanned in the Z direction.

3. RESOLUTION CUBE

The spatial image resolution is presented by the number of pixels "m" and "n" corresponding to the scanned area in the X and Y direction, while range information Z is presented by the "l" number of gray levels, creating the 'resolution cube' shown in Fig. 3. We introduce a new metric, volume of resolution (VOR) of 3-D range images. The resolution volume is described as the multiplication of the number of resolution elements in x and y and "l," the number of gray levels. The resolution volume is described by resolution units called *voxels* (volume element) corresponding to one unit in each of direction.

4. RESOLUTION AND QUANTUM NOISE CONSIDERATION

In any imaging laser radar system there is a limit to the bandwidth of the electronics. These limitations apply to both types of systems operating either with a two-dimensional array or a single detector with external scanning. This could be explained as follow: In the first type, using time-of-flight principle, acquiring the 3-D image occurs in one shot. However, the display of this 3-D image requires converting the 3-D information into serial information. This conversion essentially is a form of scanning. In contrast, systems that use a single detector and an external optical scanner, scanning occurs prior to acquiring the data. Regardless, if the scanning occurs prior or after acquiring the 3-D image, the bandwidth limitation should be the same. The bandwidth limitation has three effects: (1) limits on the total number of resolution units in 3-D images (e.g volume of resolution) (2) causes resolution ambiguity between range and spatial resolutions. (3) limits the minimum resolvable range. On the other hand increased bandwidth also leads to increased noise in the system.

5. RANGE AND RANGE RESOLUTION

To resolve the difference between two different ranges on a target, the change in the signal ΔS as a result of range change ΔR (as in Fig. 4), for the first order approximation can be calculated by differentiating S with respect to the range R, multiplying this term by the noise ratio yields the differential signal to noise ratio.
The differential signal to noise may then be calculated as:

$$(\Delta S)NR = \frac{\Delta S}{N} = \frac{4\left(P_o G e\eta / h\nu\right)^2 \left(\frac{2\omega}{c}\right) sin\left(2\omega R_0 / c\right)^2 \Delta R}{3e^2 c\left(P_o G cos(2\omega R_0/c) + P_B\right)\eta\Delta\nu + 2ei_d\Delta\nu + (4kT_e\Delta\nu / R_l)} \qquad (2)$$

The denominator contains all the noise terms[5] associated with the photodetector. For a more exact treatment we should also include the noise contribution of the optical amplifier[6]. This will be carried out in later work. The first term in the denominator includes the shot noise contribution and the correlated background noise contribution due to the background optical power P_B. Note that P_B will be much larger in the case of intentional jamming, since most of the noise power will be within the system bandwidth. The second term in the denominator is the detector dark current noise power, while the last term is the thermal noise. Here i_d is the detector dark current, R_l is again the detector load resistance, T_e is the effective temperature, and $\Delta\nu$ is the system electrical bandwidth which is directly related to the scan rate of the LADAR. For a scanning LADAR, τ_p, the time per pixel is equal to time-on-target τ, divided by the total number of pixels NP, and if "m" and "n" are the numbers of pixels in the x and y direction respectively, we have:

$$\tau_p = \tau / NP = \tau / mn \tag{3}$$

According to sampling theory, the minimum required bandwidth to represent the number of pixels should be the inverse of the time-on-target multiplied by half:

$$\Delta v = \frac{1}{2\tau_p} = \frac{mn}{2\tau} \tag{4}$$

6. SPATIAL RESOLUTION AND RANGE AMBIGUITY

If we define the quantity g as:

$$g = \frac{8\tau(P_o G e\eta/h\nu)^2\left(\frac{2\omega}{c}\right)cos(2\omega R_o/c)^2}{3e^2c(\,P_o G\,sin(2\omega R_o/c)+P_B\,)\eta+2ei_d+(\,4kT_e/R_l\,)} \tag{5}$$

being a factor that is proportional to the modulation frequency, laser power, gain and inversely proportional to the system noise.

Then, by substitution in equation 2, we obtain:

$$\Delta R = \frac{mn}{g}(\Delta S)NR \tag{6}$$

We thus have a situation where there is an inverse correlation or ambiguity between the range and spatial resolutions. Because it always essential to have (ΔS) NR>1 in order differentiate between two ranges

$$mn\frac{1}{\Delta R} < g \tag{7}$$

We can now define a minimum detectable range or noise equivalent differential range NE(ΔR), analogous to the noise equivalent ΔT in passive thermal imaging systems,[7] given by:

$$NE(\Delta R) = mn\left(\frac{1}{g}\right)\left[(\Delta S)NNR = 1\right] = mn\left(\frac{1}{g}\right) \tag{8}$$

7. RESOLUTION VOLUME AND DENSITY

The output of an ideal mixer (phase detector) for phase comparison has, as one may easily see in Fig. 2, a limited range over which the response is substantially linear.

The range interval, as has been shown, that is available to be scanned is inversely proportional to the modulation frequency, while the effective scanned range is related to the square waveform duty cycle factor *d*. It is possible to rearrange equation 6 into the following form:

$$mn\frac{(R_1 - R_2)d}{\Delta R} = \frac{2\pi c}{8\omega}\frac{gd}{(\Delta S)NR} = \frac{2\pi c g_0 d}{(\Delta S)NR} \tag{9}$$

where we multiplied both sides of the identity by $(R_1 - R_2)d$ and substituted $g_0\omega$ for g. The $(R_1\text{-}R_2)d/\Delta R$ term equals the effective number of range pixels.
Accordingly, the Volume of Resolution (VOR) under quantum noise limit is

$$VOR = mnl = \frac{2\pi c}{8\omega}\frac{g}{(\Delta S)NR} = \frac{\pi c g_0}{4(\Delta S)NR} \qquad (10)$$

where the differential signal-to-noise ratio, although mentioned earlier as being greater than 1, will ultimately be defined by the user to reflect what is an acceptable detection vs. false alarm ratio.

8. QUANTUM RESOLUTION DENSITY

If one proceeds to scan and synthesize the 3-D image for the range interval between R1 and R2 and over the area X x Y, then the physically scanned volume is given by multiplying *mnl*.

$$V = \Omega_s R^2 (R_1 - R_2)d = \frac{2\pi c \Omega_s R^2 d}{8\omega} \qquad (11)$$

where Ω is the solid angle.
We now introduce two new quantities, the *quantum resolution density,* which is defined as the ratio between the quantum resolution volume to the physically scanned volume:

$$\text{Quantum resolution density} = \frac{mnl}{V} = \frac{g}{\Omega_s R^2 d(\Delta S)NR} = \frac{g_0\omega}{\Omega_s R^2 d(\Delta S)NR} \qquad (12)$$

This parameter or metric has significance when we compare the maximum resolution achievable from detector noise considerations with that set by system parameters such as A/D converter resolution.

9. SCANNING AND GEOMETRICAL OPTICS CONSIDERATIONS

For optimal performance in terms of spatial resolution, the laser beam should illuminate exactly one pixel and adjacent pixels just touch. Normally, one would think that this would be the preferred mode of operation. However, as we pointed out in the previous section about trade-off range and spatial resolution, one may operate in under scan mode depending whether we want maximum spatial resolution or maximum range resolution.

If A_S is the total illuminated area (image frame area), and Ω_S is the subtended solid angle, as shown in Fig. 5, we thus obtain from simple geometrical considerations that the maximum number of pixels (MNP) which corresponds to the best achievable spatial resolution is given by:

$$MNP = \frac{A_s}{W_t^2} = \frac{\Omega_s}{\theta_T^2} = \frac{\Omega_s R^2}{(W_0 + 2R\theta_T)^2} \qquad (13)$$

where, as shown in Fig. 6, W_0 is the initial width, and W_t is the width of the LADAR beam at the target, and R is the range. θ_T is the half divergence angle.

10. SYSTEM RESOLUTION

In systems, the resolution volume has two other constraints: (a) The maximum number of spatial resolution points on the image (b) the maximum number of range resolution levels determined by the Analog to Digital converter resolution. For example in a k-bit Analog-to-Digital-Converter, the number the number of range resolution levels is given by 2^k. Accordingly, the system resolution volume is the number of voxels (volume elements) multiplied by the resolution points determined by the A/D converter.

The *system resolution density* then is the ratio between the system resolution volume and the physically scanned volume, and is now given by:

$$\text{system resolution density} = \frac{mnl}{V} = \frac{g}{\Omega_s R^2(\Delta S)NR} \tag{14}$$

The system resolution volume should never exceed the *quantum resolution volume*. That is:

$$\frac{2^k \Omega_s R^2}{(W_o + 2R\theta_T)^2} \le \frac{2\pi c}{8\omega} \frac{g}{(\Delta S)NR} = \frac{\pi c g_0}{4(\Delta S)NR} \tag{15}$$

where we have derived the system resolution volume from the MNP in equation 13 and compared it with the quantum VOR in equation 10. If the system resolution volume becomes larger than the quantum resolution, this means that the system has been over designed and is operating outside of a meaningful operating region. It is essential to reduce either the number of scanning points on the target or to reduce the resolution of the A/D conversion otherwise aliasing is happening. However, if the system resolution volume becomes smaller than the quantum resolution, then there should be room for improvement. Either increasing the number of scanning points on the target, or increasing the resolution of the A/D conversion can achieve this.

11. RESOLUTION EFFICIENCY

We can now define a new parameter, the *resolution efficiency* of the LADAR system, which is given by the ratio of the system resolution density and the predicted optimal resolution density.

Thus:

$$\eta_r = \text{Resolution efficiency} = \frac{\text{System resolution density}}{\text{Optimal resolution density}} = \frac{\text{System resolution volume}}{\text{Optimal resolution volume}} \tag{16}$$

This constitutes a figure of merit for a given system, in terms of achievable range and cross-range resolution. The resolution efficiency always should be less than 1, optimal resolution is achieved for η_r=1.

12. CONCLUSIONS

We have carried out a detailed analysis of range resolution and image plane spatial resolution issues in imaging laser radars. We also introduce a new gating method for determining range that combines the high resolution of phase comparison techniques with design aspects of time-of-flight measurements to provide optimal linear range measurement. We have shown that for a scanning single detector laser radar, the image plane resolution is inversely related to the range resolution, based on purely signal to noise ratio considerations. Here the limiting factor is the quantum noise in the optical detection system. We find that the noise equivalent differential range NE(ΔR) is proportional to the number of pixels in the LADAR cross range image. We introduce a new metric, volume of resolution (VOR), and we show again, from quantum noise considerations, that there is a maximum resolution volume, that can be achieved, for a given set of system parameters. Thus increased range resolution may be obtained at the cost of decreased image resolution. In a scanning single detector system, since there are no inherent restrictions to the number of pixels, the scan rate and the acquisition rate may be varied at will to control the effective image resolution. Our analysis shows that this should also be effective in optimizing the range resolution of the LADAR system. Thus an adaptive scanning system will allow on-the-fly resolution tradeoffs to be made in practical LADAR. This could conceivably do even on a frame-by-frame basis. We also analyze the effect of the analog-to-digital converter that digitizes the range receiver output. We introduce resolution efficiency, η_r, as a new figure of merit for LADAR. This parameter describes essentially how well a system performs, in terms of resolution; under a specific set of operating parameters as compared to its optimal achievable resolution performance derived from purely quantum noise considerations. We analyze how the resolution efficiency could be utilized to improve the resolution performance of a LADAR system. We also introduce in this context, the concepts of resolution volumes and resolution densities for LADAR. While our analysis has been derived from the consideration of a scanning imaging LADAR system, the results obtained should be valid for all LADAR systems, including staring array flash LADAR systems.

REFERENCES

1. E. A. Wachter and W. G. Fisher, " Coherent-burst laser ranging: decoupling resolution and unambigous range," Opt. Lett. 22(8), 570-572 (1997).
2. B.L. Stann, A. Abou-Auf, K. Aliberti, M. Giza, G. Ovrebo, W. Ruff, D. Simon, M. Stead, "Research progress on a focal plane array ladar system using a laser diode transmitter and FM/cw radar principles," Laser Radar Technology and Appplications VII, Gary W. Kamerman, ed., Proceedings of SPIE Vol. 4723, 19-30(2002).
3. B.L. Stann, W. C. Ruff, Z. G. Sztankay, "Intensity–modulated diode laser radar using frequency modulation/continuous wave ranging techniques," Optical Engineering 35(11), 3270-3278 (1996).
4. Introduction to Radar Systems, M. I. Skolnik, McGraw-Hill, 1962.
5. J. A. Overbeck, M. S. Salisbury, M. B. Mark and E. A. Watson, "Required energy for a laser radar system incorporating a fiber amplifier or an avalanche photodiode," Appl. Opt. 34(33), 7724-7730 (1995).
6. L. Morvan, D. Dolfi and J.-P. Huignard, "Optically pre-amplified LIDAR-RADAR," Laser Radar Technology and Appplications VI, Gary W. Kamerman, ed., Proceedings of SPIE Vol. 4377,284-293(2001).
7. Infrared Engineering, R. D. Hudson, John Wiley, 1969.

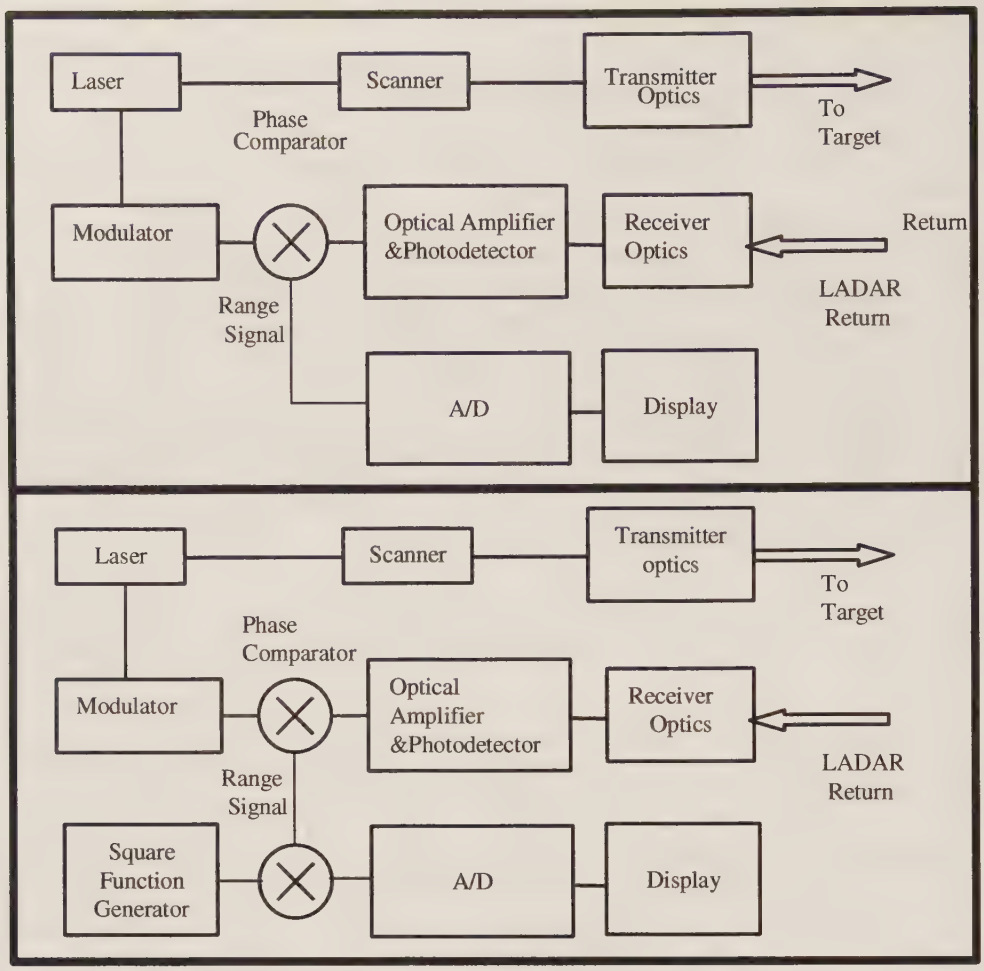


Fig. 1. Block diagrams of a phase comparison LADAR system (A) Is the block diagram of the conventional laser radar (B) Modified version of (A) that combines both phase comparison and time-of- flight.

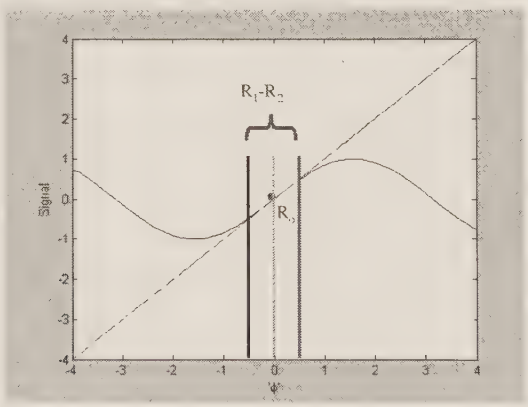


Fig. 2. The output of the phase comparison based laser radar system as a function $\Phi=2\omega R/C$, ω is the modulation frequency and R is the range.

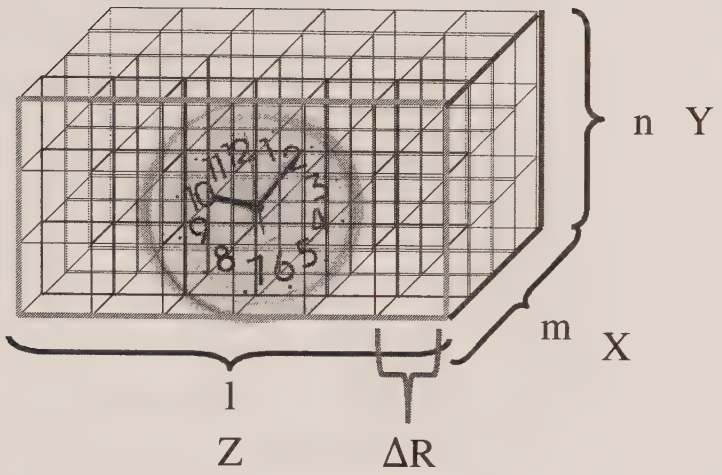


Fig. 3. The resolution cube for presenting 3-D LADAR images.

Fig. 4. Illustration to resolve between signals for two different ranges.

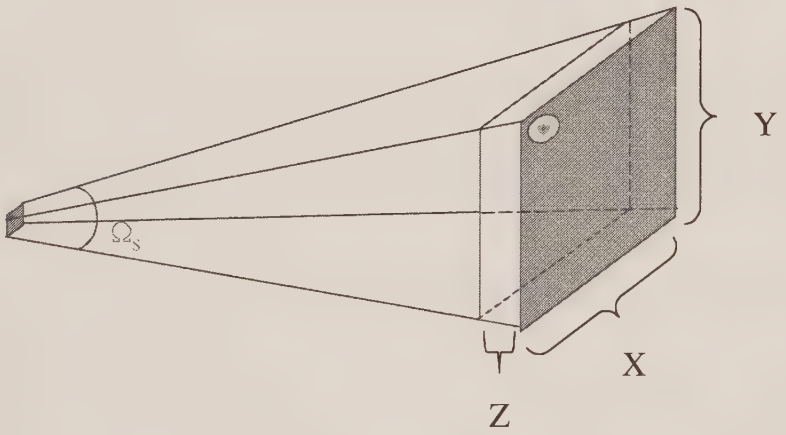


Fig. 5. Illustration of scanned volume in terms of range and solid angle.

Fig. 6. Illustration of laser beam divergence along its way from the transmitter to the target.

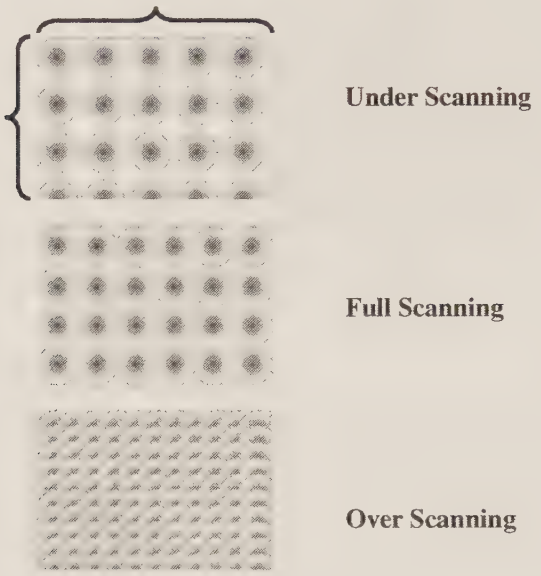


Fig. 7. Scanning modes of the target (a) under scanning (b) full scanning (c) over scanning.

Hyperspectral Imaging using Electro-Optic Fourier Transform Spectrometer

Tien-Hsin Chao, Hanying Zhou,
Jet Propulsion Laboratory
4800 Oak Grove Drive, Pasadena CA, 91109

Xiaowei Xia, Steve Serati,
Boulder Nonlinear Systems Inc.
450 Courtney Way, #107, Lafayette, CO 80026

Key Words: Hyperspectral Imaging, solid-state Imaging Fourier transform spectrometer, birefringent phase retarder, liquid crystal achromatic phase switch

ABSTRACT

JPL and BNS Inc. are jointly developing a compact, low mass, Electro-Optic Imaging Fourier Transform Spectrometer (E-O IFTS) for hyperspectral imaging applications. The spectral region of this spectrometer will be 1 – 2.5 μm (1000 – 4000 cm^{-1}) to allow high-resolution, high-speed hyperspectral imaging applications. The specific applications for NASA's missions will focus on the measurement of a large number of different atmospheric gases simultaneously in the same airmass. Due to the use of a combination of birefringent phase retarders and multiple achromatic phase switches to achieve phase delay, this spectrometer is capable of hyperspectral measurements similar to that of the conventional Fourier transform spectrometer but without any moving parts. In this paper, the principle of operations, system architecture and recent experimental progress will be presented.

1. INTRODUCTION

Traditional Fourier transform spectrometers possess two major advantages over grating, prism, and circular variable filter (CVF) spectrometers. One is the time-multiplexing effect. The Michelson interferometer's single detector views all the wavelengths (within the sensor passband) simultaneously throughout the entire measurement. This effectively lets the detector collect data on each wavelength for the entire measurement time, measuring more photons and therefore, results in higher signal-to-noise ratio, at best for situations where the source is stable. The other is the throughout advantages since the FTS does not need spatial filters (e.g. slit) in the optical light path.

However, Traditional FTIR spectrometers, used in space flight missions, obtain their optical delay by physically translating one or more optical components. The so-called translation mechanism usually dominates the risk, cost, power consumption, and performance of such instruments because:

1) Over the course of a 5-year mission, tens of millions of strokes will be required, making wear or fatigue a serious risk

2) The moving optical element cannot be rigidly held, making it sensitive to vibration and requiring that it be "caged" during launch to prevent damage, adding risk (failure of the caging mechanism to reopen).

3) Accelerating and decelerating the optical elements can torque the spacecraft, making it difficult to maintain accurate pointing.

A high-resolution FTIR spectrometer without moving parts therefore represents a substantial improvement in reliability, mission duration, and performance. It also promises to be much smaller in size and mass.

Optical Pattern Recognition XV, edited by David P. Casasent, Tien-Hsin Chao, Proc. of SPIE Vol. 5437 (SPIE, Bellingham, WA, 2004) · 0277-786X/04/$15 · doi: 10.1117/12.548075

2. ELECTRO-OPTIC IMAGING FOURIER TRANSFORM SPECTROMETER

2.1 PRINCIPLE OF OPERATION AND SYSTEM ARCHITECTURE

This paper describes the design of a novel, high-throughput, liquid-crystal-based IFTS for operation in the infrared spectral band (1-5). This approach takes advantage of fast switching ferroelectric or dual-frequency nematic LC modulators to quickly vary the time delay in order to implement fast-acquisition of the time-series data. This on-going effort is an extension of the successful development experience of an IFTS prototype that operated in the visible band (500-800nm) by Boulder Nonlinear Systems Inc. (BNS).

This proposal specifically addresses the need for reducing the weight, size and power consumption of Fourier transform spectrometers (FTS) by eliminating all moving parts. As shown in Figure 1a, in a traditional FTS, the critical time delay data is acquired by scanning one mirror within a Michelson interferometer. Post-processing of the output time measurements using Fast Fourier transform (FFT) will retrieve the spectral data. In the new E-O IFTS, the bulky, slow scanning mechanism is replaced with a solid-state time delay design using Electro-Optic (E-O) components as shown in Figure 1b. After the input beams passing through a polarizer and a linear retarder made of a birefringent material, the speed difference in refraction indexes between the ordinary and the extraordinary components will generate a time delay in the two throughput beams.

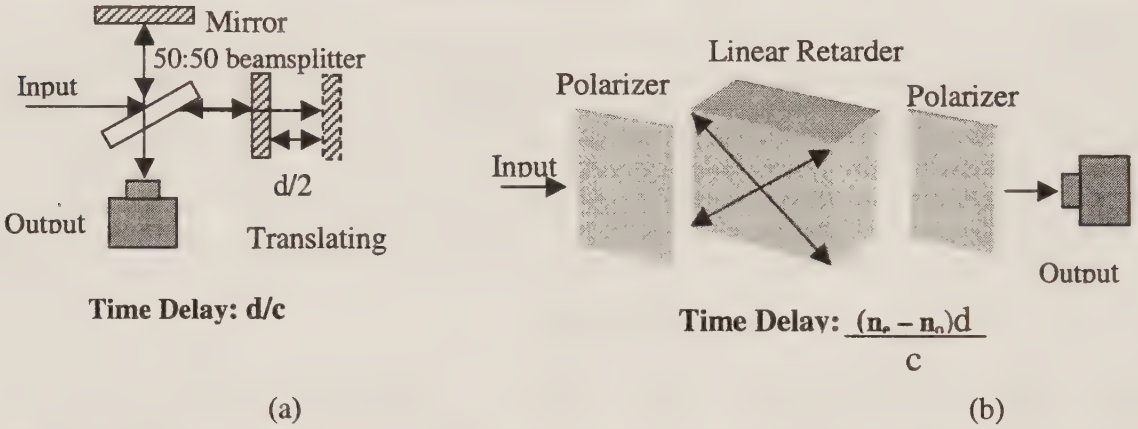


Figure 1. Time delay mechanisms of (a) a Michelson interferometer based FTS; and (b) an E-O solid-state FTS.

As shown in Figure 2, the E-O FT spectrometer would be built upon a sequence of the time-delay unit. It consists of an input polarizer, a passive quarter-wave plates (phase shifter), a series of N liquid crystal based electro-optic switches (S_1, S_2, ..S_n) interlaced with a series of (N+1) passive birefringent wave retarders (Γ_1, Γ_2, .. Γ_{n+1}), as described in figure 1a, and an output polarizer. The basic building block of system is the unit consisting of a single achromatic half-wave switch sandwiched between two neighboring passive wave retarders. The principle is that one can select between the sum or difference in total retardation of the wave passing through these two passive wave retarders by rotating the in-between achromatic half-wave switch. With parallel passive retarders oriented at 45 degrees to the input polarization, an achromatic half-wave retarder oriented at 0-degree gives the difference in retardation, while an orientation of 45 degrees gives the sum. By stacking (n+1) passive retarders, with (n) achromatic half-wave switches between them, the 2^n states of the structure represent all combinations of sums and differences of the retardance values. By using a geometric relationship of passive retarder thickness, an arithmetic progression in time delay steps is achieved.

The output of the FTS is a periodic representation of the original bandlimited input spectrum. This periodicity results from the fact that the autocorrelation of the field is sampled to recover the spectrum. Due to the limited number of time-samples, the output spectrum is more accurately a smoothed periodic representation of the input. Knowing this, one can consider the input as a single cycle of the resulting periodic output spectrum. Since the output spectrum is periodic, we can of course express it as a Fourier series

$$S(\omega) = A_0 + \sum_{m=1}^{\infty} A_m \cos(\frac{2\pi m\omega}{\Delta\omega}) + \sum_{m=1}^{\infty} B_m \sin(\frac{2\pi m\omega}{\Delta\omega}) \quad (1)$$

Where $\Delta\omega$ is the bandwidth of the spectrum and the Fourier coefficients are given below

$$A_0 = \frac{1}{\Delta\omega}\int_{-\frac{\Delta\omega}{2}}^{\frac{\Delta\omega}{2}} S(\omega)d\omega \quad (2a)$$

$$A_m = \frac{2}{\Delta\omega}\int_{-\frac{\Delta\omega}{2}}^{\frac{\Delta\omega}{2}} S(\omega)\cos(\frac{2\pi m\omega}{\Delta\omega})d\omega \quad (2b)$$

$$B_m = \frac{2}{\Delta\omega}\int_{-\frac{\Delta\omega}{2}}^{\frac{\Delta\omega}{2}} S(\omega)\sin(\frac{2\pi m\omega}{\Delta\omega})d\omega \quad (2c)$$

The total power on the detector is the integral of the input spectrum modulated by the transmission function of the FTS

$$P = \int_0^{\infty} S(\omega)T(\omega)d\omega = \frac{1}{2}\int_0^{\infty} S(\omega)d\omega + \frac{1}{2}\int_0^{\infty} S(\omega)\cos(\frac{2\pi m}{\Delta\omega}\omega + \varphi)d\omega \quad (3)$$

Where we have used the transmission function of the FTS that contains active LC switches and multi-order retarders between two polarizers,

$$T = \cos^2(\frac{\Gamma}{2} + \frac{\varphi}{2}), \quad (4)$$

Where $\Gamma = \frac{\Delta n d}{c}\omega$ is the phase retardation of the multi-order retarder for one stage, determined by its thickness, *d*, and birefringence, $\Delta n = (n_e - n_0)$. *c* is the speed of light. φ is a pure phase shift, as provided by an achromatic polarization switch. As described in the above, for multiple stages, the retardation can be stepped in increments of the fundamental τ_0 according to

$$\Gamma = m\tau_0\omega = \frac{2\pi m}{\Delta\omega}\omega \quad (5)$$

Assume that the achromatic phase shifter φ switches between four states separated by a quarter-wave of retardation (i.e. φ = $-\frac{\pi}{2}$, 0, $\frac{\pi}{2}$ and π), which was done by using the combination of an achromatic half-wave switch (S_{n+1} in Fig. 1) with an achromatic quarter-wave switch (λ/4 in Fig. 1). For one particular multi-order retardance (m), we generate four power measurements on the detector $P_m^0, P_m^{\pi}, P_m^{\frac{\pi}{2}} and P_m^{\frac{-\pi}{2}}$. Then, in the window of bandwidth, we have

$$P_m^A = P_m^0 - P_m^{\pi} = \int_0^{\infty} S(\omega)\cos(\frac{2\pi m\omega}{\Delta\omega})d\omega = \int_{\omega_0-\frac{\Delta\omega}{2}}^{\omega_0+\frac{\Delta\omega}{2}} S(\omega)\cos(\frac{2\pi m\omega}{\Delta\omega})d\omega \quad (6a)$$

$$P_m^B = P_m^{\frac{\pi}{2}} - P_m^{\frac{-\pi}{2}} = \int_0^{\infty} S(\omega)\sin(\frac{2\pi m\omega}{\Delta\omega})d\omega = \int_{\omega_0-\frac{\Delta\omega}{2}}^{\omega_0+\frac{\Delta\omega}{2}} S(\omega)\sin(\frac{2\pi m\omega}{\Delta\omega})d\omega \quad (6b)$$

$$P_0 = P_m^0 + P_m^{\pi} = P_m^{\frac{\pi}{2}} + P_m^{\frac{-\pi}{2}} = \int_0^{\infty} S(\omega)d\omega = \int_{\omega_0-\frac{\Delta\omega}{2}}^{\omega_0+\frac{\Delta\omega}{2}} S(\omega)d\omega \qquad (6c)$$

Compared with Eq. (1), the spectrum is constructed by assembling the Fourier series using the sequence of power measurements

$$S(\omega) = \frac{1}{\Delta\omega}\left[P_0 + 2\sum_{m=1}^{\infty} P_m^A \cos(\frac{2\pi m\omega}{\Delta\omega}) + 2\sum_{m=1}^{\infty} P_m^B \sin(\frac{2\pi m\omega}{\Delta\omega})\right] \qquad (7)$$

This shows that the FTS can, in principle, recover a spectrum. The Fourier coefficients are measured using structures between the polarizers: (1) A digital switch that changes the period of a sinusoidal transmission function in multiples of the input spectral bandwidth, and (2) A four-state achromatic switchable retarder that steps in quarter-wave increments at each order.

The common-path design is physically compact and inherently stable with no moving parts. By replacing precision mechanical movements with low-power binary liquid crystal switches, power supply and packaging requirements are greatly reduced which translates into a large weight reduction. In addition, the common-path approach simplifies the optical design. The in-line optical design reduces size and its mechanical stability and durability. These benefits allow the optical head to be fabricated as a solid block optic with no gaps between elements.

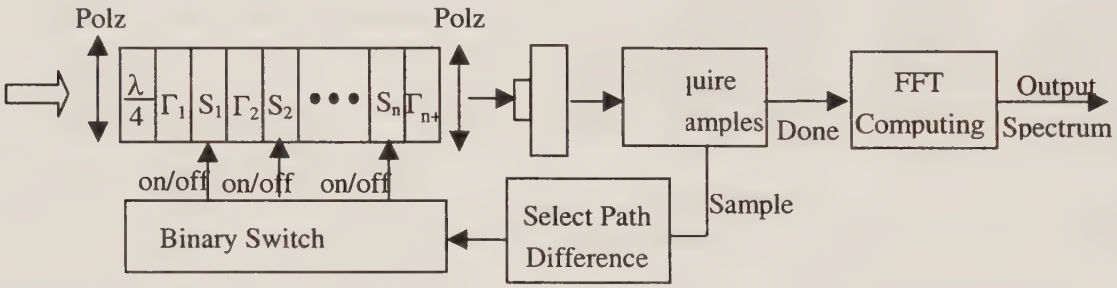


Figure 2. Block diagram of the electro-optic Imaging Fourier transform spectrometer

2.1 SPECTRAL RESOLUTION

Similar to a conventional FTS, the spectral resolution, $\Delta\sigma$, of the proposed EO-FTS is related to the maximum optical path difference, Δx, or equivalently, the maximum time delay, δ_{max}, between the two interfering waves: $\Delta\sigma = \frac{1}{\Delta x_{max}} = \frac{1}{d\cdot(n_o - n_e)} = \frac{1}{\lambda\cdot 2^N}$. If a total of N switches (N stages) is used, the time delay of each switch will be approximately $2^0\lambda$, $2^1\lambda$,..., $2^N\lambda$ with maximum time delay $\delta_{max} \sim 2^N\lambda_{med}$ where λ_{med} is the central wavelength of the spectral passband). For the proposed spectrum range of 1.0 ~ 2.5μm, the central wavelength is about 1.8 μm (5500 cm^{-1}). Thus the spectral resolution for an 11-stages and a 13-stages EO-FTS will be about 2.68cm^{-1} and 0.67cm^{-1} respectively.

3. DEVELOPMENT OF A 3-STAGE IR FTS

We have recently developed a 3-stage IRFTS and demonstrated preliminary IR spectral data capture and retrieval. Details of this development work including the fabrication and testing of achromatic half-wave FLC switches, optical design, electronics and software development will be presented. The three-stage IR FTS successfully identified the 1310 nm Diode laser spectrum.

3.1 OPTICAL DESIGN AND SETUP

The setup is sketched in Figure 3a. A Diode laser at 1310 nm wavelength was used as light source and a single silica photodiode (*New Focus*, Model#2011) as the light detector. The signal from the detector was acquired by an A/D board in a computer. Multiple optical delays are generated via three achromatic half-wave FLC switches (S1, S2 and S3) put among four passive wave retarders with delays of 1λ, 2λ, 4λ and 8λrespectively.The fourth half-wave switch (S4) was used to generate a 0 or π phase retardation. An achromatic $\lambda/4$ switch was used to shift the system to additional $\pm\pi/2$ phase. An interface control program was written to apply voltages (±5 V) on the FLC switches, and to acquire signals. Figure 3b shows a photo for the Lab setup of the spectrometer.

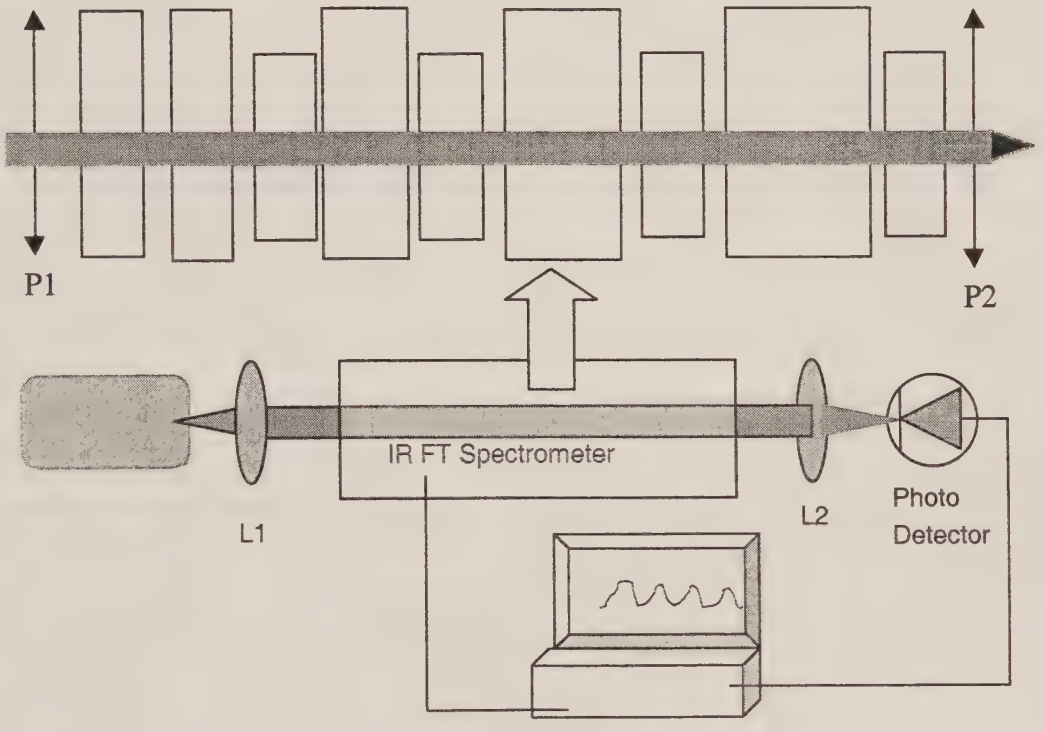


Figure 3a. Schematic setup for the 3-stage spectrometer test.

Figure 3b. Photo of Lab setup for the 3-stage spectrometer.

3.2 FABRICATION OF ACHROMATIC HALF-WAVE SWITCHES

The schematic of the achromatic half-wave solid-state switch is shown in Figure 4. Two passive Nitto Denko retardation film (NRF) sheets and one FLC switch were used to construct the compound retarder. The retardance for each NRF is a full-wave at 600 nm. The retardance of the FLC switch is a half-wave at 1120nm. Four achromatic switches have been fabricated.

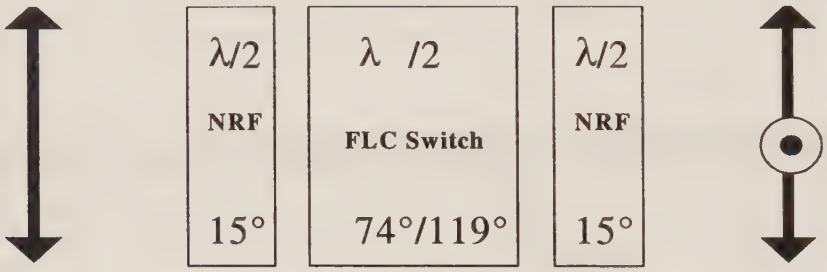


Figure 4. The Schematic of an achromatic half-wave switch

A typical test result (green solid curve) is shown in Figure 5, compared with the simulation results (red dotted curve). During experimental measurements, the LC switch was toggled at a rate of 0.4 Hz, with a driving voltage of ±4.8V. The switch speed of the modulator can be fast, on the order of 100 μs, with low operating voltages (<5V). The spectral range with leakage less than 0.5% is from 0.93 μm to 1.43 μm. The simulation parameters used are: full-wave retardance for each NRF at 600 nm, half-wave retardance of the FLC switch at 1220 nm, and the orientations of the three waveplates at 15°, 73.7°, and 15°, respectively. It can be seen that the experiments matches the simulation well.

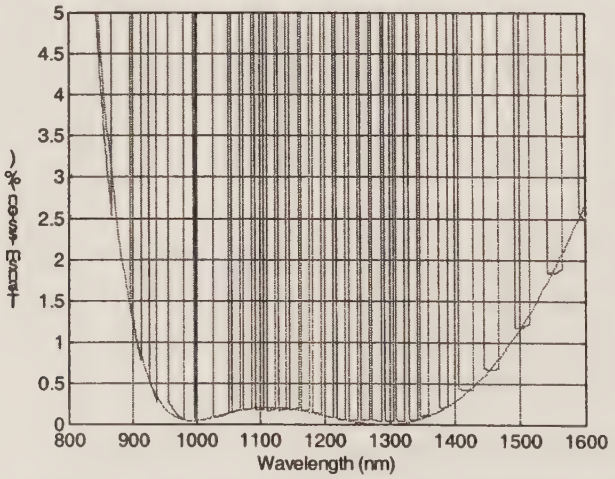


Fig.5. Theoretical and experimental comparison of the achromatic Half-wave switch

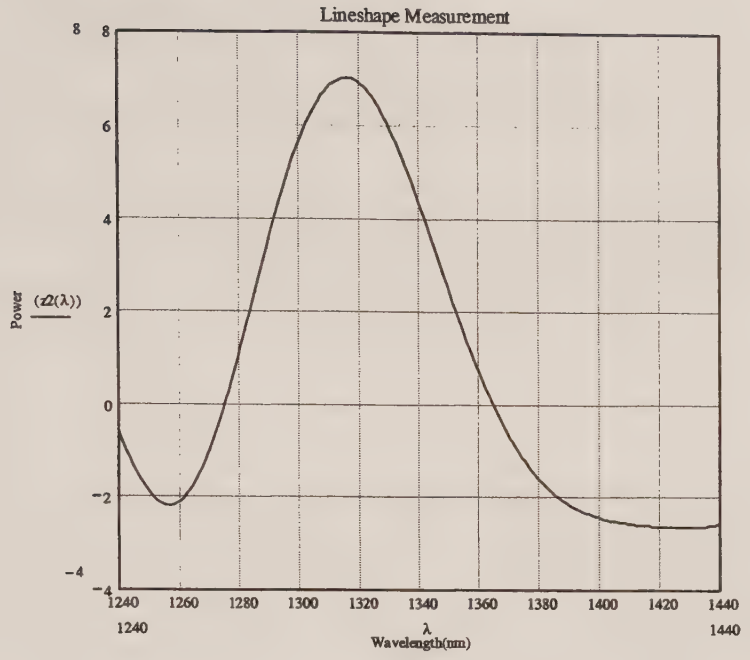


Figure 6. Measured and recovered spectrum of a laser source centered at 1310 nm.

We have developed a data recover algorithm using MathCAD. The algorithm has the operation for Fourier transformation, incorporated the dispersion effect of the quartz which was measured experimentally. The recovered spectrum identifying the 1310-nm laser wavelength is shown in Figure 6. The broadened lineshape is due to the rough resolution of the 3-stage system (~145nm width).

4. Summary

We have developed principle of operation and system architecture of a electro-optic Fourier transform spectrometer. Initially, we have demonstrated a proof-of-principle 3-stage IR FTS breadboard. The major components design, fabrication and test results have been reported. The experimental result of spectral data capture of that of a test diode laser source and the follow-on recovery through an inverse Fourier transform have also been accomplished. Future work includes the development phase retarders with of high birefringence material (e.g. YVO_4) to greatly reduce the system light path is underway. We will also develop a multiple stage system with ultrahigh resolution. Imaging optics will also be included into the system to enable hyperspectral imaging applications

5. ACKNOWLEDGMENTS

The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. The authors wish to express their sincere thanks to Dr. Geoffrey helpful technical discussions.

6. REFERENCES

[1] G. D. Sharp *et al.* "Polarization Interferometer Using Multi-order and Zero-order Birefringence Switches", U.S. Patent Application 08/948,860, Claims allowed (10/12/99).

[2] G. D. Sharp, P. Wang, S. Serati, T. Ewing, "Liquid Crystal Fourier Transform Spectrometer", SPIE AeroSense, (1998).

[3] P. Fellgett, Cambridge University thesis, 1951.

[4] G. D. Sharp, K. M. Johnson, "Liquid crystal achromatic compound retarder" , U. S. Patent 5,658,490

[5] Xiaowei Xia, Steve Serati, Teresa Ewing, Ping Wang, "A massively parallel liquid crystal Fourier transform spectrometer", Final report of Air Force Phase-II SBIR project, Dec. 1999.

Feature Selection from High-dimensional Hyperspectral and Polarimetric Data for Target Detection

Xue-wen Chen[a] and David Casasent[b]

[a]Information and Telecommunication Technology Center, Department of Electrical Engineering & Computer Science, The University of Kansas, Lawrence, KS 66045

[b]Department of Electrical & Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213

ABSTRACT

Hyperspectral and polarimetric data contain spectral response information that provides detailed descriptions of an object. These new sensor data are useful in automatic target recognition applications. However, such high-dimensional data introduce problems due to the curse of dimensionality, the need to reduce the number of features used to accommodate realistic small training set sizes, and the need to employ discriminatory features and still achieve good generalization (comparable training and test set performance). In this paper, we evaluate both hyperspectral and polarimetric feature sets and identify features useful for distinguishing targets from background. Various feature selection algorithms are assessed in terms of the goodness of the selected features and computation time. Our results show that (1) the integration of branch and bound algorithm and floating forward selection algorithm is promising for hyperspectral and polarimetric target detection applications; and (2) the combination of both hyperspectral and polarimetric features yields significantly better classification results than either hyperspectral or polarimetric features alone.

Key words: branch and bound, feature selection, floating forward selection, hyperspectral and polarimetric data, target detection.

1. INTRODUCTION

Hyperspectral (HS) and polarimetric data are very powerful and provide excellent detailed information useful for many target detection applications. This is evident as initial HS results have achieved improved detection of objects in clutter [1-3] and polarization imaging research has shown that manmade objects are sources of emitted and reflected polarization while backgrounds such as vegetation and soil are predominantly unpolarized [4]. Thus, using both hyperspectral and polarimetric images may provide significant improvement on the detection of targets from backgrounds and the discrimination of background from manmade targets. Hyperspectral and polarimetric image data processing is quite new in military vehicle detection. Thus, new hyperspectral and polarimetric algorithms for military target detection applications are needed.

One of major challenges in the use of HS and polarimetric data is the high dimensionality of the input data available (a large number of features N_F) and the limited number of training samples. When the number of training samples N_i (per class i) does not satisfy [5] $N_i > 10N_F$, then *generalization* problems are expected [6], i.e, training and test set classification performance P_c differ. The training set size is rarely adequate and thus this issue is of concern in HS and polarimetric data processing. This causes poor estimates of matrices such as the covariance matrix C and its inverse C^{-1} required in many discrimination algorithms. In order to mitigate these problems, data dimensionality reduction is generally required. Thus, we address feature selection algorithms in this paper.

Optical Pattern Recognition XV, edited by David P. Casasent, Tien-Hsin Chao, Proc. of SPIE Vol. 5437 (SPIE, Bellingham, WA, 2004) · 0277-786X/04/$15 · doi: 10.1117/12.541414

Feature selection is to select a subset of m features from a larger set of n features to optimize the value of a criterion function J over all subsets of the size m. There are many different feature selection algorithms used in the literature. Sequential forward selection (SFS) and sequential backward selection (SBS) [7] are two widely used sequential feature selection methods. The SFS method first selects the best single feature and then adds one feature at a time which in combination with the selected features maximizes the criterion function J; the SBS method starts with all input features and successively deletes one feature at a time. Both SFS and SBS methods are computationally attractive. By dynamically controlling the number of forward and backtracking steps, floating sequential search methods [8] have been shown to perform better than both SFS and SBS algorithms. However, all these sequential methods are not guaranteed to produce a feature subset which yields the global maximum value of the criterion among all possible subsets, since "the best pair of features need not contain the best single feature" [5]. The only optimal feature selection algorithms are exhaustive search and branch and bound (BB) [9]. In an exhaustive search, the best solution is found by evaluating the criterion function J over all possible combinations of features, which is impractical or impossible, since the number of combinations of features increases exponentially as the dimensionality increases. The other optimal feature selection method which avoids an exhaustive search is the BB algorithm. This algorithm is efficient because it avoids an exhaustive search of the whole search space by rejecting many subsets which are guaranteed to be suboptimal and it guarantees that the selected subset is the global optimal solution for a given criterion function. The BB algorithm works for many practical problems for which an exhaustive search would be impractical or impossible. More efficient BB algorithms [10, 11] are recently proposed and are shown to outperform the original BB algorithm by reducing redundant J evaluations in a BB algorithm. A two step feature selection algorithm with sequential forward selection as a filter and branch and bound algorithm as the final feature selector is also proposed for high dimensional feature selection applications in which even the most efficient BB algorithms alone are not practical because of the huge amount of computations [12].

The objectives of this paper are two: (1) to evaluate both hyperspectral and polarimetric features and to select best feature sets for target recognition applications; and (2) to assess the performance of various feature selection algorithms (including forward selection, floating forward selection, and branch and bound based algorithms) in terms of computation time and the goodness of selected features. In addition, a two step feature selection algorithm based on floating forward selection and branch and bound algorithm is proposed and evaluated. Our experimental results show that *the combination of HS and polarimetric data is capable of distinguishing targets from backgrounds and outperforms HS data alone.*

The paper is organized into four sections. Section II describes feature selection algorithms. Section III gives the experimental results. Finally, conclusions are drawn in Section IV.

2. FEATURE SELECTION METHODS

2.1 Forward Selection

Forward selection is the simplest greedy search algorithm. It starts from an empty set and continues to add one feature at a time, which, when combined with the previous selected features, yields the best results. The algorithm terminates when the desired number of features is reached.

2.2 Floating Forward Selection

Floating forward selection (FFS) algorithm combines ideas from both forward selection and backward selection and introduces a dynamic backtracking mechanism to avoid the nesting problem. Like FS algorithms, FFS algorithm starts from an empty set and selects the best feature using forward selection strategy; after each forward selection, it continues to perform backward reevaluation steps as long as the values of a selected criterion function increases. This backtracking strategy aims to remove the worst features which are considered as the best features in previous steps. FFS algorithms avoid the nesting problems presented in most FS algorithms, as features selected can be later discarded.

2.3 Branch and Bound Algorithm

Branch and bound algorithm is an optimal searching method, which guarantees the optimal solution and avoids exhaustive search in the whole feature space. In the BB algorithm, a search tree is produced with n - m levels to select m features out of n, and one feature is omitted at each level of the tree. The algorithm first follows one complete path of n - m nodes and computes an estimate (a bound B) on the criteria function J being used. J is then evaluated at the nodes at a given level of the tree; if $J < B$ for a given node, then J need not be evaluated at nodes under that node (J decreases as we proceed down the tree, since fewer features are then used, i.e., more features are omitted). To further reduce the searching space, some modifications include [11]: efficient evaluation of paths through the tree with no branches, starting searches at levels 1/4 and 1/2 of the way down the tree, ordering the features (nodes) by which are best using forward selection, and calculation of the initial bound B using the m best features selected by forward selection. At the upper levels of the search tree, we do not expect $J < B$, since only one or two features (the best ones by forward selection) are omitted. Thus, starting searches further down the tree speeds-up the search (as J must be evaluated for most of the nodes at the top of the tree). Ordering the features by some estimate of which features are best (the original BB algorithm uses an arbitrary order) significantly speeds up the search since the best features are at the top of tree and omitting these "better" features first is expected to produce a $J < B$ sooner (hence J need not be evaluated for nodes below such a node). Rapid selection of a good initial large B (using features ordered by any estimate of which are best) also makes $J < B$ more likely to occur earlier in the tree search (thus, allowing evaluation of J to be omitted for regions of the tree).

2.4 Two-step Branch and Bound Algorithm

Branch and bound algorithm is efficient compared to exhaustive search algorithms and has been shown useful in practical applications [11]. However, when applied to high dimensionality problems (e.g., select 10 features out of 100 original features), BB or its modified versions are slow. In these cases, a two-step search method is preferred: the first step reduces the feature size (filtering) and the second step selects the best subset of features from the remaining features. In [12], a forward selection is first applied to reduce the original feature size to about 30; a modified branch and bound algorithm [11] then selects the best subset of features from the 30 features. In this paper, we propose to use floating forward selection to select an initial subset of 28 features (more are possible, as the number of subsequent calculations is not excessive). We then use the modified branch and bound algorithm to select the best m of n features. This is still an optimal search algorithm, given the reduced set of initial features. It is faster than BB and much faster than an exhaustive search. In addition, as we will demonstrate, floating selecting algorithms perform better than forward selection algorithms and avoid nesting problems exhibited in forward selection algorithms.

3. EXPERIMENTAL RESULTS

3.1 Hyperspectral and Polarimetric Data

The dataset used contains images with 480 by 640 pixels taken at spectral range 460 – 1000 nm with 20 nm step at 4 different polarizations (0°, 45°, 90°, and 135°). Each scene contains one military vehicle in a vegetation background. The objective is to locate the military vehicles in the scenes. We use a portion of the target and background pixels in the hyperspectral and polarimetric set of vehicle images (e.g., HMMWV, personnel carrier, etc.) as the training set and the test set of *pixel data*. Detection test results (*image/scene test results*) are then presented on the vehicle images which are not used in the training stage. We refer to these as *test scenes*. The image result for the scene from which the training and test set pixel data were extracted is referred to as the *training scene*. From the training scene, we randomly extract 200 pixels belonging to the target and 400 pixels belonging to the background. These samples are divided into two equally sized groups, one as training sample sets and the other as test set. In this paper, we report results on these training and test pixel data.

3.2 Comparison of Feature Selection Algorithms

We first compare different feature selection algorithms: forward selection (FS), floating forward selection (FFS), standard branch and bound (BB), modified branch and bound (MBB). The criterion function used to evaluate the selected feature sets is Mahanaobis distance J. All four selection algorithms were run on the full set of n = 28 spectral features with zero polarization. Table 1 summarizes J results for the four algorithms. Both standard BB and the modified BB algorithms gave the highest J values in all cases. The J value increases as the number of features increases. FS produced lowest J values. FFS algorithm yielded slightly lower J values than the optimal solution. The computation times for the four algorithms were summarized in Table 2. As seen, FS was the fastest; FFS was faster than both BB and the modified BB algorithms. Our results also showed that the modified BB was faster than BB. When the number of original features is manageable (28 in this case), the modified BB is preferred as it guarantees to find optimal solutions in an efficient manner. However, if the number of original features is large (e.g., 50), BB based algorithms are slow [11]. In that case, suboptimal solutions are needed such as FFS. To search in a very high dimensional feature space, we propose to use FFS algorithm to select 28 initial features and the modified BB is applied to these 28 features.

Table 1: Comparison of feature selection algorithms: Mahanaobis distance

Number of selected features	Forward selection	Floating forward selection	Standard branch and bound	Modified branch and bound
1	0.0059	0.0059	0.0059	0.0059
2	0.0260	0.0260	0.0309	0.0309
3	0.0312	0.0312	0.0437	0.0437
4	0.0418	0.0418	0.0547	0.0547
5	0.0589	0.0595	0.0595	0.0595
6	0.0614	0.0630	0.0630	0.0630
7	0.0643	0.0644	0.0644	0.0644
8	0.0660	0.0660	0.0660	0.0660
9	0.0671	0.0671	0.0675	0.0675
10	0.0686	0.0686	0.0686	0.0686
11	0.0696	0.0696	0.0701	0.0701
12	0.0706	0.0706	0.0713	0.0713

Table 2: Comparison of feature selection algorithms: computation time (second)

Number of selected features	Forward selection	Floating forward selection	Standard branch and bound	Modified branch and bound
1	0.0100	0.0200	0.2510	0.1000
2	0.0200	0.0100	0.5910	0.1100
3	0.0200	0.0100	3.1740	0.4310
4	0.0200	0.0200	2.4140	0.6910
5	0.0300	0.0600	1.7330	0.4000
6	0.0300	0.0700	0.7820	0.2800
7	0.0400	0.0800	1.6830	0.5000
8	0.0500	0.0900	3.6150	1.0220
9	0.0600	0.1000	5.1080	1.7920
10	0.0700	0.1200	5.8590	2.1130
11	0.0700	0.1300	5.2170	2.5740
12	0.0800	0.1400	4.2660	1.5630

3.3 Comparison of Feature Sets using a two step branch and bound algorithm

In this section, we compare HS features, polarimtric features, and the combination of HS and polarimetric features. We first consider three different feature sets: the 28 HS features with 90 degree polarization; the 56 polarization ratio features, and the 56 polarization difference features, extracted from both hyperspectral images and polarimetric signatures. Let the intensity distributions obtained at orthogonal polarizations with the wavelength λ be represented by $I_{\|}(x, y, \lambda)$ and $I_{\perp}(x, y, \lambda)$ (e.g., 0 versus 90 degrees and 45 versus 135 degrees), where *(x, y)* are the positions on the image. For each wavelength λ, we will exploit the polarization differences to create polarization-different images (PDI) $I_{\|}(x, y, \lambda) - I_{\perp}(x, y, \lambda)$, and the polarization ratios to create polarization-ratio images (PRI) $\frac{I_{\|}(x,y,\lambda)}{I_{\perp}(x,y,\lambda)}$. Thus, for the hyperspectral images taken at spectral range 460 – 1000 nm with 20 nm step at 4 different polarizations (0°, 45°, 90°, and 135°), we create 56 polarization-difference images (PDI) (28 PDIs from HS images with 0° and 90° polarizations and 28 PDIs from HS images with 45° and 135° polarizations) and 56 polarization-ratio images (PRI). These features contain information from both hyperspectral images and polarimetric signatures. Thus, they are very attractive to target detection and discrimination. A portion of the target and background pixels in polarization-difference images and in polarization-ratio images will be used as our training and test sets of pixel data. Comparison between polarization difference images and polarization-ratio images will be conducted to decide which images will be used for the military vehicle detection applications.

From the previous section, we demonstrated that the modified BB guarantees optimal solutions while is very efficient if the number of original features is less than 30. To compare different features sets, we use the modified BB if the number of features is 28, and the FSS followed by the modified BB if the number of features is greater than 28. We use the same Mahanaobis distance as our criterion function. To see how the selected features perform in test data, we use k-nearest neighbor with k = 3 as the classifier, as our ultimate goal is for data classification. Table 3 shows the classification results for both training and test data for 28 spectral band features (with the 90 degree polarization). Table 4 lists the results for polarization difference images and Table 5 for polarization ratio images. As seen, in all cases, the generalization is good for the number of selected features range between 1 and 10. Among the three sets of features, HS with 90 degree polarization yields slightly better performance than PDI features in terms of classification accuracies, while PRI yields the lowest classification accuracies.

We combine both the original 28 HS features with 90 degree polarization and the 56 polarization difference images and select a subset of features from the 84 features for target detection. Table 6 list the classification rates for different number of features selected. Again, the generalization is good: there is no big difference between training and test results. Thus, the selected features are expected to predict unseen data well. In addition, the classification accuracies with both HS and polarimetric features for both training and test data are greater than these with either 28 HS features or 56 PDI features alone. The highest classification rate for test data using the 84 features is 97%, which is about 6% higher than these using 28 HS features or 56 PDI features. This clearly demonstrates that *combination of both hyperspectral and polarimetric information is necessary for target detection*. Thus, with polarization information, better classification accuracies are achieved for both training and test data sets.

Table 3: Classification results using 3 nearest neighbor rules for the 28 HS features with 90 degree polarization

Number of selected features	Classification rate (training)	Classification rate (test)
1	0.8800	0.8833
2	0.8667	0.8300
3	0.9200	0.8867
4	0.9100	0.8600
5	0.9100	0.8700
6	0.9367	0.8900
7	0.9233	0.8900
8	0.9200	0.9100
9	0.9200	0.9067
10	0.9233	0.9067

Table 4: Classification results using 3 nearest neighbor rules for the 56 features extracted from polarization difference images

Number of selected features	Classification rate (training)	Classification rate (test)
1	0.6600	0.6000
2	0.8133	0.7300
3	0.8267	0.8533
4	0.8800	0.8367
5	0.8767	0.9000
6	0.8800	0.8767
7	0.8733	0.8867
8	0.8933	0.9067
9	0.8933	0.9067
10	0.8833	0.9133

Table 5: Classification results using 3 nearest neighbor rules for the 56 features extracted from polarization ratio images

Number of selected features	Classification rate (training)	Classification rate (test)
1	0.6233	0.7233
2	0.7633	0.7833
3	0.8167	0.8267
4	0.8100	0.8233
5	0.8100	0.8300
6	0.8367	0.8700
7	0.8333	0.8467
8	0.8267	0.8633
9	0.8167	0.8500
10	0.8567	0.8567

Table 6: Classification results using 3 nearest neighbor rules for the combination of 28 HS features and 56 PDI features

Number of selected features	Classification rate (training)	Classification rate (test)
1	0.6333	0.5800
2	0.8667	0.8300
3	0.9433	0.9467
4	0.9500	0.9433
5	0.9600	0.9467
6	0.9433	0.9567
7	0.9633	0.9533
8	0.9567	0.9600
9	0.9700	0.9600
10	0.9600	0.9700

4. DISCUSSION AND CONCLUSION

Automatic target detection using both hyperspectral and polarimetric features is discussed. Various feature selection algorithms are compared. Our results show that the modified branch and bound algorithm can find optimal solution in an efficient manner. Also, results indicate that floating forward selection algorithm is better than forward selection algorithm and is capable of finding features which are suboptimal. When applied in very high dimensional space, the combination of floating forward selection and the modified branch and bound algorithms is proposed and is shown to be promising. We also compare different feature sets: hyperspectral features only, polarization difference image based features, and polarization ratio image based features. Our results show that among the three feature sets, HS features with 90 degree polarization yields best classification accuracies for both training and test data sets. PDI features yield slightly lower classification accuracies than HS features. The combination of HS and PDI features, however, gives us the best results. Thus, both hyperspectral and polarimetric information is needed for target detection. Future research includes comparison of different classifiers and target detection from test image scenes.

ACKNOWLEDGE

This material is based upon work supported by the U. S. Army Research Laboratory and the U. S. Army Research Office to Xue-wen Chen under contract number DAAD19-03-1-0123. We thank Neelma Gupta of ARL for providing the HS vehicle image data used.

REFERENCES

1. B. Yu, L. Hoff, I. Reed, A. Chen, and L. Stotts, "Automatic target detection and recognition in multiband imagery: a unified ML detection and estimation approach," *IEEE Trans. Image Processing*, vol. 6, pp. 143-156, 1997.
2. L. Hoff, A. Chen, X. Yu, and E. Winter, "Enhanced classification performance from multiband infrared imagery," in *Proceeding of the IEEE Twenty-Ninth Conference on Signals, Systems and Computers*, 1995.

3. L. Hoff, J. Zeidler, and C. Yerkes, "Adaptive multispectral image processing for the detection of small targets in terrain clutter," in *Signal and Data Processing of Small Targets*, O. Drummond, ed., Proc. SPIE 1698, pp. 100-114, 1992.
4. J. Howe, M. Miller, R. Blumer, T. Petty, M. Stevens, D. Teale, and M. Smith, "Polarization sensing for target acquisition and mine detection," *Proc. SPIE*, vol. 4133, 2000.
5. A. Jain, R. Ruin, and J. Mao, "Statistical pattern recognition: a review," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, pp. 4-37, 2000.
6. E. Baum and D. Haussler, "What size net gives valid generalization?" *Neural Comput.*, vol. 1, pp. 151-160, 1989.
7. K. Fukunaga, *Introduction to Statistical Pattern Recognition,* Academic Press Inc. New York, 2nd Ed. 1992.
8. P. Pudil, J., Novovicova, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognition Letters*, 15, pp. 1119 – 1125, 1994.
9. P. Narendra and K. Fukunaga, A branch and bound algorithm for feature subset selection, IEEE Trans. on Computers, C-26, pp. 917-922, 1977.
10. B. Yu and B. Yuan. "A more efficient branch and bound algorithm for feature selection," *Pattern Recognition*, vol. 26, pp. 883-889, 2000.
11. D. Casasent and X.-W. Chen, "Mine and vehicle detection in hyperspectral data: waveband selection," *Proceedings of SPIE, Automatic Target Recognition,* Orlando, April, 2003.
12. D. Casasent and X.-w. Chen, "Aflatoxin detection in whole corn kernels using Hyperspectral methods", *Proc. SPIE*, vol. October, 2004.

Fast Non-Invasive Eyetracking and Eye-Gaze Determination for Biomedical and Remote Monitoring Applications

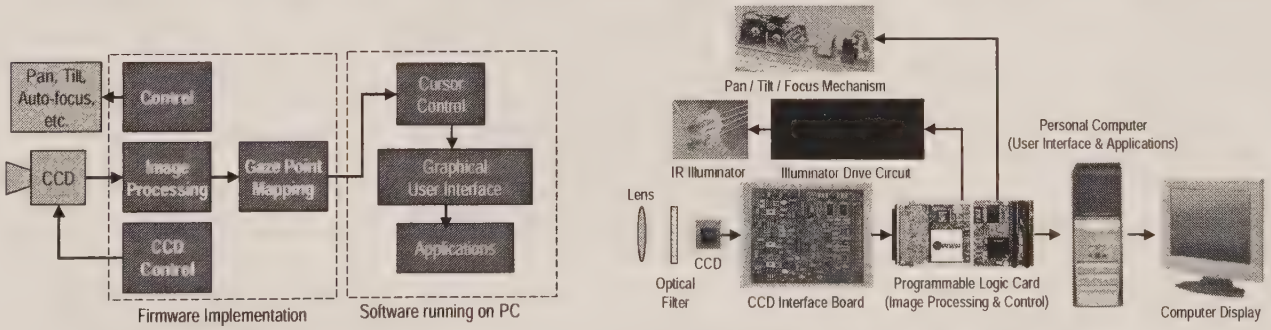
ASHIT TALUKDER, JOHN-MICHAEL MOROOKIAN, S. MONACOS, R. LAM, C. LEBAW, A. BOND
Intelligent Instruments and Systems Technology Group, In-Situ Instruments Section
Jet Propulsion Laboratory
4800 Oak Grove Drive, MS 300-123, Pasadena, CA 91109
USA
Email: {Ashit.Talukder, Johnmichael.Morookian, Steve.Monacos}@jpl.nasa.gov
URL: http://eis.jpl.nasa.gov/~atalukde

Abstract: - Eyetracking is one of the latest technologies that has shown potential in several areas including human-computer interaction for people with and without disabilities, and for noninvasive monitoring, detection, and even diagnosis of physiological and neurological problems in individuals. Current non-invasive eyetracking methods achieve a 30 Hz rate with possibly low accuracy in gaze estimation, that is insufficient for many applications. We propose a new non-invasive visual eyetracking system that is capable of operating at speeds as high as 6-12 KHz. A new CCD video camera and hardware architecture is used, and a novel fast image processing algorithm leverages specific features of the input CCD camera to yield a real-time eyetracking system. A field programmable gate array (FPGA) is used to control the CCD camera and execute the image processing operations. Initial results show the excellent performance of our system under severe head motion and low contrast conditions.

Key-Words: - Detection, CCD (Charge coupled device), Eyetracking, Human-computer Interaction (HCI), Tracking, Image Processing Hardware, Field programmable Gate Array (FPGA)

1. INTRODUCTION

Non-invasive human-computer interaction is a field that promises to ease the communications between machines and humans, by providing a easy-to-use interface on the machine that humans can understand, while



(a) System software architecture (b) System hardware architecture

Figure 1: Components in the Eyetracking system

simultaneously providing the tools that machines can use to interpret the actions of humans.

Recent technological advancements in human-computer interaction include audio-based interactions such as speech recognition [1], natural language processing and speech synthesis. Visual recognition and interface

Optical Pattern Recognition XV, edited by David P. Casasent, Tien-Hsin Chao, Proc. of SPIE Vol. 5437 (SPIE, Bellingham, WA, 2004) · 0277-786X/04/$15 · doi: 10.1117/12.548073

tools include face detection [2] and recognition [3], facial expression estimation [4], and gesture recognition of sign language with hand motion [5]. While such interface mechanisms are useful for providing information about the identity of a user, and for explicit communication between a human and a computer, they are not very useful for assisting handicapped people who have trouble speaking or moving. Eyetracking and automatic gaze point determination is a technology that can be used as an alternative interface tool to assist normal and handicapped individuals, and also provide a plethora of other physiological information for various applications, such as visual acuity determination for opthalmology applications, reading disorders, evaluation of user-interface designs, etc.

Eyetracking has traditionally been achieved using several approaches, many of which require the user to wear a device, thereby hindering their normal activities. This includes design of a head-mounted video eye movement system at Applied Science Laboratory, and high-magnification tracking using biomicroscopes. In electrooculagraphy, electrodes are placed on the temporal bones near each eye and on the forehead (nasal); eye motion causes changes in electric field potential, that can be used to estimate gaze direction. In a conventional limbus-based eye tracking method [6], the eye is illuminated by infrared light, usually coming from a light emitting diode (LED). The amount of light reflected back from the eye's surface is measured with a photodiode and this gives the position of the eye. As the cornea and iris reflect much less infrared light than the sclera does, the intensity of the reflected light is proportional to the relative amounts of sclera and cornea within the acceptance angle of the detector and hence eye position. While the limbus tracking method

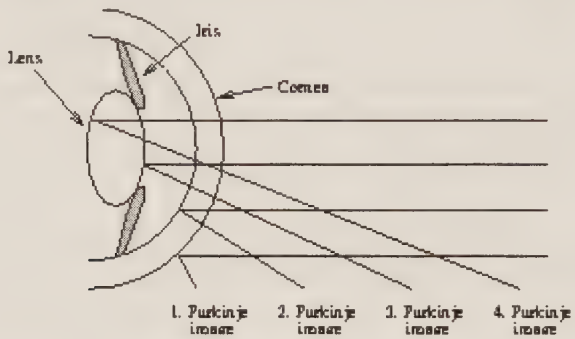


Figure 2: Reflection-based method(s) for eye tracking.

previously involved mounting the photo-diodes and LEDs directly at the eye, recent developments have used fibre optic cables to relay the infrared light to and from the eye, thereby requiring the system to be placed very close to, but not necessarily mounted on, the individual. Magnetic search coils designed by Skalar Instruments have been used where large electromagnetic coils surround the patient and induce voltage changes relative to the position of fine coils embedded in special annular contact lenses worn by the patient. However, ocular discomfort typically occurs in 20 minutes using this approach. A thorough review of prior techniques for eyetracking has been done in [7].

These prior eyetracking solutions are mostly invasive, and work at a maximum rate of 60 Hz. Non-invasive and remote gaze point determination at significantly higher rates would open the gates to a host of other applications that current eyetracking solutions cannot address.

Our solution is a non-invasive, real-time eyetracker using a fast, remote CCD video camera and associated image processing for locating eye features from a distance and estimating the gaze point. The system operates at 6-12 Khz and is ultimately expected to have head-tracking incorporated to allow eyetracking during free motion of the head. The software architecture of the system is shown in Figure 1a, and the hardware components are shown in Figure 1b.

We first discuss non-invasive eyetracking techniques in Section 1.1, followed by a discussion of potential applications (Section 1.2) that could arise from the high-speed, non-invasive eyetracking solution that we

propose. The system architecture for our eyetracker is presented in Section 2, followed by a discussion of the new CCD video camera in Section 3. Our image processing algorithms for eye feature location is discussed next (Section 4), followed by a description of our gaze mapping technique (Section 5). Initial results are presented in Section 6.

1.1. Operational Principle

Current non-invasive eyetracking solutions are principally divided between three approaches. All three solutions use the fact that the eye reflects incident light from various surfaces that can be then used to track and eye and estimate the gaze direction. When light is shined into the user's eye, several reflections occur on the boundaries of the lens and cornea, the so-called Purkinje images (Figure 2). The first Purkinje image off the corneal surface is also called the glint.

The first solution to eye gaze mapping uses the location of the glint with standard incident visible light and the dark pupil, and uses an artificial neural network to estimate gaze coordinates based on glint and pupil locations [8]. However, this approach requires extensive training examples for calibration, and its gaze mapping is not highly accurate.

The first and fourth Purkinje images can be used for tracking the direction of gaze by the Dual-Purkinje Image technique [9], which uses the relative positions of these reflections to calculate the direction. The Dual-Purkinje-Image technique is generally more accurate than the prior technique, but the main disadvantage with this technique is that the fourth Purkinje image is rather weak, so the surrounding lighting must be heavily controlled.

Our eyetracking system uses the Pupil-Center/Corneal-Reflection (PCCR) method (Figure 2) to non-invasively determine the eye's gaze direction. The main concept of the PCCR technique is to locate the pupil center and the center of the corneal reflection off the eye surface, and use these two centroids to determine the gaze direction of a user. To this end, a video camera is oriented close to the user's nominal gaze direction such that it is focused on the user's eye.

A small, low power, infrared emitting diode (IRED) located at the center of the camera lens illuminates the eye. The IRED generates the corneal reflection and causes the bright pupil effect, which enhances the camera's image of the pupil, as seen in Figure 3. This bright pupil effect (Figure 3), caused by light that enters the eye and is reflected by the fundus at the rear of the eye back through the pupil, is also commonly known as the red-eye effect that is noticed predominantly during flash photography. The aperture of the pupil acts as a sink for the IR radiation, while the iris acts as a reflector. An IR sensitive video camera can therefore be used to

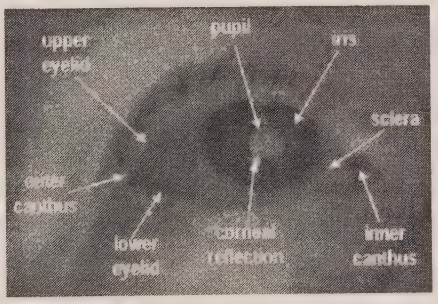


Figure 3: IR illuminated image

obtain a high contrast black and white image of the pupil (black) and the iris (bright).

For many human/machine interface applications, we are interested in the point at which the gaze direction intersects a computer display rather than the gaze direction itself. Thus, we employ a gaze mapping algorithm which maps the relationship between the pupil and corneal reflection centroids to the user's gaze point on the computer display. The relative locations of the pupil and the corneal reflection vary systematically as a

function of the direction of gaze. The vector difference between the pupil and the CR can be correlated to objects that are fixated.

1.2. Eyetracking Applications

Development of a high-frame rate eyetracking system enables a broader class of applications than current commercially available systems. For example, refractive surgery requires accurate positioning of the laser beam onto the center of the pupil. The incidence of high-pulse-rate lasers, such as the Wavelight system that fires 200 pulses per second, have necessitated eyetracking at speeds of 4Khz. Therefore, off-the-shelf videocamera systems that operate at 30 Hz are insufficient for refractive surgery.

Certain eye defects and physiological disorders such as amblyopia and those involving visual attention (or lack of it) can be diagnosed if knowledge of the pupil motion can be made available. Diagnosis of such disorders can be improved if the eyetracking system can operate at speeds up in the kilo-hertz range.

Several severe clinical conditions can be associated with specific types of sacaddic eye movement abnormalities. Parkinson's disease, that results in impaired neural inhibition, is characterized by sacaddic eye abnormalities such as [10] hypometria, increase latency, and reduced peak velocity. Multiple sclerosis is characterized by progressive degeneration of white matter of the brain and spinal cord, and introduces sacaddic problems [11] such as dysmetria, increased latency, duration. Myasthenia Gravis, an autonimmune disease, produces [12] dysmetria, variable waveforms, and increased duration without decreased peak velocity, with normal latency. Therefore, high speed accurate eyetracking and eye motion analysis could result in diagnosis of such diseases. A good discussion of eye motion and related diseases is detailed in [13].

It has also been shown in previous work [14, 15] that vital neurological signs such as heart-rate , and inter-cranial pressure can be extracted by evaluating pupil size, shape, equality, and it's response to light, and charting the changes in these pupil characteristics over time. Therefore, extremely high-speed eyetracking is necessary in order to achieve desired improvements in biomedicine such as refractive surgery and medical diagnosis.

2. SYSTEM ARCHITECTURE

Many commercial eyetracking systems that operate using the PCCR technique rely on standard video cameras, which produce produce full-frame data at approximately 30 frames-per-second. The video stream is typically processed in a computer using either the Central Processing Unit (CPU) or a Digital Signal Processor (DSP) on a frame-grabber board. Such systems are limited to slow-speed, full-frame operation and place undue requirements on the host computer system.

In contrast, our eyetracking system implements both control and image processing in reprogrammable "firmware" on a Field-Programmable Gate Array (FPGA) (Figure 1b and 4). Because the time-intensive processing is implemented in the FPGA, the burden is removed from the host computer. The firmware developed for the FPGA is relatively easily extended to a full-custom design in an Application Specific Integrated Circuit (ASIC), for applications in which small system size or mass production are desired. We envision the natural progression of this architecture to a standalone computer peripheral similar to an optical mouse.

We also implement a novel control mechanism for a commercial charge coupled device (CCD) imager, which makes region-of-interest (ROI), or sub-window, readout possible at high-rates. Since the features of interest (the corneal reflection and pupil) typically occupy a small subsection of the entire camera frame, ROI capability allows the user's gaze direction to be determined at high rates.

An infrared illuminator is chosen so that the illumination will not be distracting to the user and because the red-eye effect is most pronounced in the near infrared. To minimize motion-blur effects, we incorporate a control circuit that pulses the IRED on and for short durations during the image acquisition period. The CCD imager is preceded by an optical bandpass filter that is tuned to pass the IRED light while rejecting ambient illumination.

Control of the CCD image acquisition is implemented in the FPGA firmware. When in acquisition mode and searching for the pupil and corneal reflection in an image, the camera operates in full-frame mode. After these features have been located, the camera drops into a fast tracking mode in which only pixels in a subwindow around the pupil position in the previous frame are read. The subwindows may be redefined on a frame-by-frame basis, allowing the pupil and corneal reflection to be tracked as they move from frame to frame. Greater motion of the user can be accomodated with the addition of a camera pan/tilt/zoom system, currently under development. Control for this system would also be implemented in the FPGA.

The image processing algorithm, implemented in the FPGA, reports the pupil and corneal reflection x and y pixel coordinates for each input frame. This results in a huge reduction of data, from 350,000 bytes of image data to just a few bytes of centroid coordinates. The resulting centroid data can easily be sent across a low-bandwidth link such as USB to be processed by the host CPU to determine the user's gaze point on the computer display. We have chosen to use Transtech's PMC-FPGA01 card as our FPGA development platform because its architecture can support high-speed image transfer to a computer via PCI bus for debugging and development purposes.

3. HIGH-FRAME RATE CCD CAMERA

A real-time eye-tracking system requires an image capturing device that operates at an extremely high frame rate in order to be able to effectively track the pupil in the presence of saccadic motion. Therefore, a high-frame rate image capture device is necessary in order to build a general eyetracker system for various human-computer interaction applications.

Commercially available cameras are well suited to applications requiring full-frame, large form factor video streams or sub-window readout for still-frame image capture. They are not designed for the combination of single frame and high-speed streaming video with sub-window image capture and greater than eight bits of resolution per pixel. Development of a custom camera fills this void. We discuss a custom camera capable of region-of-interest (i.e. sub-window) read out rates up to several Kilo-Hertz with greater than eight bits of

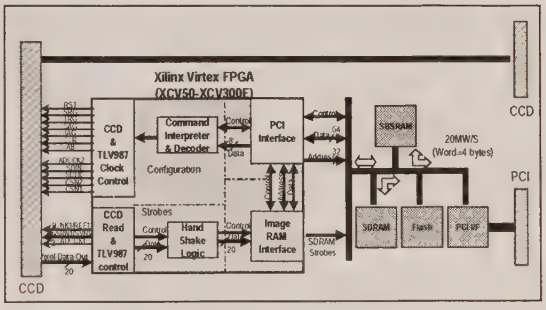


Figure 4: FPGA for custom CCD control and image processing

resolution per pixel [1].

The CCD camera consists of the FPGA interface card and TC237 CCD card as shown in Fig. 1b. This organization allows for optimizing the CCD card while maintaining a flexible architecture for communicating with the camera. The CCD card includes regulator circuitry, level translators and buffers, a TC237 imager and two TLV987 signal processors to provide TTL compatibility for all input and output signals on the card. The TC237 CCD sensor and TLV987 signal processors are the primary components of the CCD card.

The TC237 is a CCD imager with an image area of 680 pixels per line and 500 lines. The active image area of the CCD is 658 pixels per line and 496 lines and contains all exposed pixels during scene capture [3]. There are additionally 22 pixels per line and four additional lines of dark pixels, which make up the remainder of the image area and are used for background level calibration [3]. A significant point to be noted is that two lines of pixels can be read out simultaneously by performing two shift register load operations before

beginning serial read out [3], [4]. We will use this capability for more robust image processing to detect the eye features, as discussed later in Section 4.

Real-time sub-windowing capabilities in the TI CCD provides the speedup of image capture in the kilohertz range that makes it ideal for eyetracking. Read out of a sub-window consists of three fundamental operations. They include a refresh operation, scrolling through the image until the sub-window is reached and reading the pixels which are in the sub-window. The time required to read out a sub-window is defined as the sum of the refresh time, the scroll time and the window read time.

Theoretical measures about the operational speeds of the CCD camera for specified subwindow sizes can be obtained. Assuming a 10 MHz imager system clock and an additional 300usec delay for a line request, the measured CCD row read time is 14.5usec. This includes 1usec for a line request, 1usec for read out of pixels from the CCD and an additional 12.5usec to transfer the pixels across the global bus interface logic into the FPGA memory. The maximum achievable frame rate is 6KHz for this case for 8x8 sized-subwindows.

The operational characteristics of the CCD camera for the desired speeds therefore drive and limit the choice of image processing algorithms that can be used to detect the eye features (pupil and corneal centers). While a variety of image processing solutions exist for circle/ellipse finding, our image processing solution should work in tandem with the line-scanning mechanism of the CCD to ensure the fast eyetracking performance that is desired for the various applications mentioned earlier. We discuss our image processing solutions next.

4. IMAGE PROCESSING ALGORITHMS FOR FAST DETECTION AND TRACKING

The image processing algorithm for locating the pupil and cornea in the eye is a very important component of our eyetracking system. Several options exist for detecting the pupil and cornea. These include automatic thresholding followed by simple blob coloring and binary shape-based blob filtering to remove blobs that are not "circular" in shape. Such a solution that uses absolute gray-level measures, while very fast in practice, is not expected to work well when multiple reflections off the sclera are present, and when the person's skin tone is light, and therefore similar to the pupil reflection. Additionally, this solution requires that the CCD camera capture the entire image prior to processing, and does not take advantage of the fact that eyetracking time can be dramatically reduced by using the fact that subwindowing and scanning specific rows in the CCD camera is significantly faster than scanning the entire image frame.

The Hough transform has been actively used to robustly locate specific shapes in an image [16] such as lines and circles. Initial results of the Hough transform for locating circular shapes such as the pupil and cornea under different motion and contrast conditions were very promising. However, the Hough transform for detecting circles is not suited for running in real-time and it also requires the entire image frame to be acquired prior to processing.

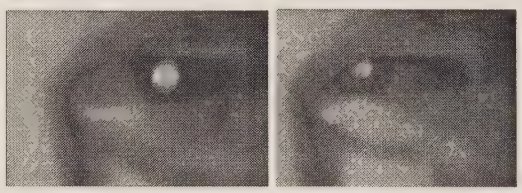


Figure 5: Typical images

We use a new slice-based image processing technique that leverages on the CCD's fast acquisition rate of image row slices, thereby ensuring a highly optimized end to end eyetracking system. We take advantage of the symmetry of the circles and the gray-scale contrast of the pupil and cornea from other parts of the eye in designing our image processing algorithm. In this approach, we determine horizontal image slices (Figure 6) that contain the pupil and cornea, and note the start and end coordinates of the pupil/cornea on each valid slice. Information from multiple slices is then combined to robustly locate the pupil and corneal centers. The

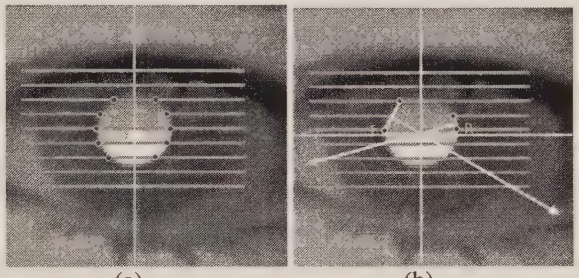
number of slices is fewer than the image row size, i.e. the slices are a subset of the image rows, as shown by the horizontal red lines in Figure 6.

We first discuss the slice-based algorithm to locate borders of a pupil in a scanned row (slice). For a given image row (slice), we apply the 1D Canny edge filter [17] to extract the high-frequency components in the scan-line. The Canny edge filter has been shown [17] to provide excellent signal-to-noise ratio in extracting edges in signals. The gray value at each pixel fused with its edge magnitude is used to mask out pixels in the slice that are not likely candidates for the pupil/corneal edges. The magnitude and sign of the edge value at each remaining pixel is then used to select likely candidates for the start and end of the pupil within that slice. Grouping of such candidate pairs based on edge magnitude, sign, and distance between pairs of pixels combined with prior knowledge of the expected size of a pupil is used to determine the pupil border pixel locations within the row (slice).

A similar method is used to detect corneal borders in a given image row slice. The gray-scale and edge threshold values used to mask candidate corneal border pixels are different from those used to locate pupil borders.

The prior steps yield a list of slices that contain pupil and corneal regions. This list might contain false alarms: slices that are incorrectly categorized as containing pupil or cornea. Additionally, some slices containing pupil/cornea may not have been detected. Therefore, we postprocess the data where we fill gaps in the slice list by grouping adjacent slices that have similar pupil/corneal midpoints along the x-axis (horizontal direction). Slices that have outlier midpoint values along the x-axis, or are isolated and do not have adjacent slices (along the y-axis or vertical direction) are rejected as false alarms. The average midpoint between the pupil edges in the remaining slices gives the horizontal (x-axis) center of the pupil, as shown by the vertical green line in Figure 6a.

In order to locate the y-center of the pupil, we choose two pairs of points on the edge of the pupil from the list of slices, compute the normal vectors to these line pairs, as shown in Figure 6b, and note the intersection of



(a) (b)

Figure 6: Horizontal slicing algorithm for locating pupil and cornea

these normals with the horizontal coordinate of the pupil center (vertical line in Figures 6a and b). The average y-coordinate of the intersection computed from several line pairs (4-tuple of points along the pupil borders) yields the y-coordinate of the pupil center.

A similar mechanism is used to locate the center of the cornea. The post processing is effectively able to detect the corneal slices in the presence of reflections on the sclera that look similar to the corneal reflection.

This approach of processing slices (image rows) makes effective use of the camera hardware. One parameter to choose is the number of rows to skip between two adjacent slices. The image processing algorithm is faster when the rows skipped are greater, at the cost of reduced resolution and risk of missing much (or all) of the pupil, especially when the eyelids cover the pupil. Conversely, pupil detection accuracy will improve greatly if few rows are skipped between adjacent slices, but at the cost of slower performance.

5. GAZE-POINT MAPPING ALGORITHM

An estimate of the gaze point (or direction) can be obtained by locating the centers of the pupil and cornea. Traditionally, gaze mapping in the PCCR uses the assumptions that: the pupil is approximately centered around the optical axis of the eye, and the corneal surface is spherical, and that light reflects off the corneal surface, i.e. the first Purkinje reflection (Figure 2) is visible. Therefore, the gaze angles (horizontal and vertical), relative to the camera pointing direction Z, can be measured by computing the vector between the pupil and corneal reflections (Figure 7). When the person looks directly at the camera, the two centers overlap,

Figure 7: Gaze point estimation concept

when the person looks to his/her right, the pupil center moves to the left of the corneal center.

If the linear difference vector between the pupil center (p = [px py]) and corneal center (c = [cx cy]) are d = [dx dy], a linear approximation to gaze point mapping can assume that:

$$\text{Xscreen} = a11 + a12\ dx$$
$$\text{Yscreen} = a21 + a22\ dy \qquad (1)$$

where Xscreen is the x-coordinate of the gaze point on the screen. This concept is illustrated in Figure 7.

In practice, a quadratic mapping is traditionally used to compensate for roll angle misalignments between the camera's and screen's horizontal axes, as shown in Eq (2).

$$\text{Xscreen} = a11 + a12\ dx + a13\ dy + a14\ dx\ dy$$
$$\text{Yscreen} = a21 + a22\ dy + a23\ dx + a24\ dy\ dy \qquad (2)$$

However, we note that the pupil may not lie in the center of the optical axis, nonlinearities can occur due to corneal flattening, and there might be misalignments and nonlinearities associated with the camera pointing vector, and the change in eye (pupil and corneal) shape as the gaze direction changes,. Therefore, we use the actual x,y image coordinates of the pupil and cornea to estimate the gaze point. We use the quadratic form in Eq (2), with the true pupil and corneal centers. This is expected to yield more general solutions than those that use only the difference vector.

6. RESULTS

Our eyegaze system is comprised of two main components: the hardware and system software. Both components are currently being developed in parallel and results of each component operating independently are available. In the system hardware, the two novel subcomponents are the CCD camera, and the FPGA for driving the camera and doing the real-time image processing. The novel software features are the image processing and the gaze-point mapping algorithms. We discuss our initial results from our system hardware and software components next.

6.1. Hardware Results

Preliminary results indicate that the camera can read 8x8 sub-windows at a maximum frame rate of 6 KHz. This rate is for 10 MHz pixel and line transfer rates. The 8x8 sub-window is positioned in the second line and column of the CCD FOV.

Tables 1 and 2 illustrates two scenarios for camera operation and the achievable frame rates for these cases. The integration time is as defined above. All three cases assume only one sub-window within the CCD FOV. The beginning of the sub-window is located at the second line and second column in the frame and is eight pixels across by eight pixels high. The imager system clock, which defines both the parallel line transfer rate and the serial pixel rate of the CCD, is 10 MHz or 15 MHz.

The first scenario is based on measured performance of the camera in the testbed. A 10 MHz imager system clock is used for this case. As can be seen in table 1, the measured row read time is 14.5usec, which includes 1usec for a line request, 1usec for read out of pixels from the CCD and an additional 12.5usec to transfer the pixels across the global bus interface logic into FPGA memory. The maximum achievable frame rate is 6KHz for this case.

The second scenario is targeted for release 2.0 of the camera and assumes a 15MHz imager system clock with simultaneous serial pixel data from both outputs of the TC237 CCD. The line request, scroll, row read and window read times include the over head for the legacy testbed as measured in the second scenario. These timings are smaller compared to the first scenario because of the higher clock speed and dual pixel readout for this case. As Table 2 shows, this scenario can achieve greater than 11 KHz frame rates with a 50% increase in the system clock.

Table 1: TC237 Maximum Frame Rates for One 8x8 Sub-Window

Readout Operation	Release 1 10 MHz	Release 2 15 MHz
Refresh Time (μS)	50	33.3
First Window Scroll Time (μS)	1	0.8
Row Read Time First Window (μS)	14.5	13.9
Window Read time First Window (μS)	116	55.7
Frame time (μS)	167	89.8
Frames per Second	5988	11135

Table 2: TC237 Minimum Frame Rates for One 8x8 Sub-Window

Readout Operation	Release 1 10 MHz	Release 2 15 MHz
Refresh Time (μS)	50	33.3
First Window Scroll Time (μS)	245	187.8
Row Read Time First Window (μS)	38.9	30.2
Window Read time First Window (μS)	311.2	120.8
Frame time (μS)	606.2	342
Frames per Second	1649	2923

6.2. Software Algorithm Results

6.2.1. Image Processing Results

The image processing algorithm is one of the critical components of our eyetracking system. Initial tests of our image processing algorithms were done on a PC platform with prototyping in MATLAB. Tests on three subjects were carried out. In two of these tests, the subject was "compliant" and no sudden head movements were carried out. The third case involved rapid head motion by the subject on purpose to test the limits and robustness of our image processing algorithms.

The first two tests were comprised of the subject looking at 20 gaze points on the screen. The last one involved random eye and head movement. All tests were done at a 30 Hz frame-rate with a total time of 20 seconds per subject. The average accuracy in detecting the pupil and centroid centers was determined by counting the number of frames where the error in locating the pupil and centroid centers is less than 15 pixels. Since no truth data is available, the true pupil and corneal centers were marked manually.

We first evaluated the image processing performance on the two subjects without artificial head motion. The average accuracy in pupil detection was 96.4%. The average accuracy in pupil detection was 96.8%. Most of the errors occurred for frames where the eyelid covered most of the pupil during the process of blinking.

For the third case where the subject intentionally moved the head rapidly, the accuracy in correctly locating the pupil and cornea was 91.4% and 92.6% respectively. The decrease in performance occurred due to deterioration in tracking performance due to rapid motion in the eye feature between successive frames. Figure 8 shows results of our image processing algorithm that accurately locates the pupil and cornea under various conditions. The red circles are the locations of the detected pupil and cornea. It locates the cornea and pupil even in the presence of strong motion blur (Figure 8b) that results in blurring and defocusing of the cornea and pupil areas. Figure 8a shows an example where only about 55-60% of the pupil is visible and 40% is covered by the eyelid. Additionally, its contrast relative to the iris is very low. The pupil center is located with high accuracy even in this case. Figure 8d shows our image processing results where 25-30% of the pupil is covered by the eyelid, and head motion causes the eye features to be blurred. The presence of reflection off the sclera further complicates the problem, but performance is robust even in this case.

The average execution time of the current prototype MATLAB image processing code is 0.7 seconds per frame. The run-time is expected to speed up tremendously after the algorithm is implemented in C/HDL in the near future. Further parallelization of the code after FPGA implementation will result in the desired real-time system for detecting and tracking the eye pupil and cornea for gaze-point mapping.

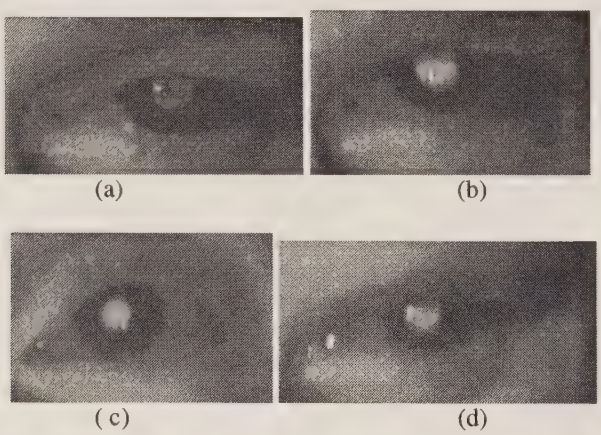


Figure 8: Results of pupil and corneal detection

6.2.2. Gaze-point Mapping Results

The centers of the pupil and cornea are used from the image processing outputs to estimate the gaze point of the eye at each frame. Considerable improvement in gaze mapping accuracy was obtained when our algorithm (that employs the actual location of the pupil and corneal centers to evaluate the gaze mapping) was used, compared to the standard approach (that employs the difference vector between the pupil and corneal centers).

During training and validation, a calibration target generator program was used to plot points (attractors) on the screen at one-second intervals, and the user was required to gaze at each point. The average of the absolute difference between the true and estimated gaze points was used as a measure of the efficacy of the gaze-mapping algorithm.

The average absolute gaze estimation error in the horizontal (x) direction was 12.3 pixels, and 30.7 pixels in the vertical direction when our gaze-mapping algorithm was used. In contrast, the standard (commonly used) gaze-mapping algorithm yielded an average absolute error of 16 pixels in X, and 40.4 pixels in the vertical direction. Further tests need to be done to evaluate and improve the performance of the gaze-mapping algorithms.

7. CONCLUSIONS AND FUTURE WORK

In this paper, we have discussed a new real-time remote eyetracking and gaze mapping algorithm that is ultimately expected to operate at 6-12 KHz, thereby making it applicable to a variety of applications that current eyetrackers cannot be used for. New image processing algorithms that take advantage of the CCD subwindowing capabilities is detailed. Initial results are promising and illustrate the efficacy of our solution for pupil/corneal detection and tracking, and it's robustness to rapid head and eye motion. The new gaze mapping algorithm is shown to yield good estimates. Further improvements to the gaze mapping algorithm are expected, using advanced nonlinear mapping such as support vector machines or neural networks. We are currently evaluating conversion of our image processing routines from the MATLAB environment to implementation on the FPGA using either SIMULINK and Real-time Workshop tools in MATLAB, or direct implementation in HDL for the FPGA.

REFERENCES

1. Rabiner, L. and B.-H. Juang., *Fundamentals of Speech Recognition.* 1993: Prentice Hall.
2. Yang, G. and T.S. Huang, *Human face detection in a complex background.* Pattern Recognition, 1994. **27**(1): p. 53 - 63.
3. Talukder, A. and D. Casasent. *Adaptive Activation Function Neural Net for Face Recognition.* in *IEEE Intl Joint Conf. on Neural Networks.* 2001: IEEE.
4. Tian, Y., T. Kanade, and J. Cohn, *Recognizing action units for facial expression analysis.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001. **23**(2): p. 97-115.
5. Starner, T., J. Weaver, and A. Pentland, *Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998. **20**(12): p. 1371-1375.
6. Kimmig, H., et al., *MR-Eyetracker: A new method for eye movement recording in functional magnetic resonance imaging (fMRI).* Experimental Brain Research, 1999. **126**: p. 443-449.
7. Collewijn, H., *Eye Movement Recording*, in *Vision research: A Practical Guide to Laboratory Methods*, C. R.H.S. and R. J.G., Editors. 1999, Oxford University Press. p. 245-285.
8. Baluja, S. and D. Pomerleau, *Non-intrusive gaze tracking using artificial neural networks.* 1994, Research Paper CMU-CS-94-102, School of Computer Science, Carnegie Mellon University: Pittsburgh PA.

9. Müller, P.U., et al., *A comparison of a new limbus tracker, corneal reflection technique, purkinje eye tracking and electro-oculography*, in *Perception and Cognition*, G. dYdewalle and J.V. Rensbergen, Editors. 1993, Elsevier Science Publishers. p. 393-401.
10. Tanyeri, S., C.J. Leuck, and T.J. Crwaford, *Vertical and horizontal saccadic eye movements in Parkinspn's disease.* Neuro-Opthalmology, 1989. **9**(165).
11. Reulen, J.H.P., E.A. Sander, and L.A.H. Hogenhuis, *Eye movement disorders in multiple sclerosis and optic neuritis.* Brain, 1983. **106**(121).
12. Yee, R.D., S.M. Whitcup, and I.M. Williams, *Saccadic eye movements in myasthenia gravis.* Opthalmology, 1987. **94**(219).
13. Ciuffreda, K.J. and B. Tannen, *Eye movement basics for the clinician*. 1995, Missouri: Mosby.
14. Yoshida, M., et al. *Statistical properties of simultaneously recorded fluctuations in pupil diameter and heart rate*. in *Proc. IEEE EMBC and CMBEC*. 1995: IEEE.
15. Calcagnini, G., et al., *Cardiovascular rhythms in spontaneous pupil diameter fluctuation.* IEEE Computers in Cardiology, 1997: p. 211-214.
16. Ballard, D.H., *Generalizing the Hough Transform to detect arbitray shapes.* Pattern Recognition, 1981. **13**: p. 111-122.
17. Canny, J., *A computational approach to edge detection.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986. **8**: p. 679-698.

Image segmentation using variable threshold functions

Hassan Hassan

Department of Electrical and Computer Engineering
Lawrence Technological University
21000 W. Ten Mile Road
Southfield, Michigan 48075-1058
Email: h.hassan@ieee.org

ABSTRACT

This paper presents digital image segmentation techniques and algorithms using variable threshold functions. The technique is successful to detect regions with different or poor light radiances and can be applied to images with occluded or noisy objects. The Variable threshold functions are derived from discrete time functions often used in digital control system design. The developed algorithms can be integrated on a single monolithic integrated circuit or implemented as an embedded system for real-time applications.

Keywords: image segmentation, threshold functions, real-time thresholding, data reconstruction, zero-order hold, first-order hold, polygonal-hold, slewer-hold, digital control, linear filters.

1. INTRODUCTION

One approach for image segmentation is feature thresholding. In feature thresholding techniques, any pixel with a feature value (e.g., gray level value) darker than a selected threshold is classified as a black point, whereas any pixel with a feature value lighter than or equal to the selected threshold is classified as a white point.

Feature thresholding techniques may be classified into two schemes: one-dimensional feature thresholding, and thresholding in multidimensional space. The one-dimensional feature threshold can be viewed as a test involving a single feature; this feature is usually the gray level value. While thresholding in multidimensional space requires two or more features such as gray level, color intensity for red, green, and blue, texture, hue, and others.

One dimensional feature thresholding can be further classified into four different approaches: global, local, dynamic, and multi-threshold global. In global thresholding, the threshold value (T) depends only on G(I,J), where G(I,J) is the gray level of the point (I,J). If T depends on both G(I,J) and N(I,J), where N(I,J) represents some local property of the point (I,J), then the thresholding approach is local. If T depends on the coordinates of the pixel (I,J) as well as on G(I,J) and N(I,J), then the thresholding approach is dynamic. In the dynamic thresholding approach, it is possible to have a unique threshold value to every pixel in the image. When more than one threshold value is selected based on G(I,J), then the thresholding approach is multi-threshold global.

More details about all thresholding paradigms are outlined in section 2. In section 3, a complete study, analysis, and variable thresholding functions derivations are given. Experimental results on variety of images using variety of variable thresholding functions are given in section 4. Final summary and conclusions are given in section 5, followed by a list of references.

2. THRESHOLDING PARADIGMS

Thresholding schemes can be classified into one of the following paradigms: global thresholding, local thresholding, multi-threshold global thresholding, dynamic or variable thresholding, and thresholding in multi-dimensional space. In the following subsections, full details about all these paradigms are given.

Optical Pattern Recognition XV, edited by David P. Casasent, Tien-Hsin Chao, Proc. of SPIE Vol. 5437 (SPIE, Bellingham, WA, 2004) · 0277-786X/04/$15 · doi: 10.1117/12.541603

2.1 Global thresholding

In the global thresholding scheme, the gray-level histogram for the entire image is usually used to determine the unique and global threshold value. A good value for the global threshold is usually selected at the bottom of the valley of the histogram.[1, 2] This selection technique was first proposed by Prewitt and Mendelsohn[1] and is known as the *mode method.* This method works well if the gray level values occupied by the object regions and the background region are sufficiently separated which results in a strong bimodal histogram.

Unfortunately, the gray level histograms for practical images are rarely bimodal. In addition, identification of valleys and peaks on the histograms is itself a very difficult problem since they may contain valleys defined by noise spikes especially when the image taken under poor lighting conditions or in noisy environments. Therefore, many researchers have tried to refine and improve the histogram to make it more bimodal to make the selection of the global threshold easier.[3-5] Other scientists have also attempted to use other information such as edge detection information to construct the gray-level histogram[6] or to aid the selection of a good threshold.[7, 8] Rosenfeld and Torre[9] have proposed histogram concavity analysis as an approach to threshold selection. They suggested that concavity points of maximum depth are good candidates for threshold value.

2.2 Local thresholding

In the local thresholding scheme, the decision whether a pixel belongs to the object region or to the background region is determined by its local gray level statistics. The gray level of a pixel is compared with the mean gray level over a specific kernel (or mask) where the pixel under consideration is at the center of the kernel.[10, 11] This method can have better performance than the global thresholding for images taken under noisy and poor lighting environments. However, it needs more computation time since the threshold for each pixel has to be determined locally.

2.3 Dynamic thresholding

In the dynamic thresholding scheme, [12, 13] local statistics are measured in overlapping windows of the image. These statistics consist of the histogram in each window and its variance. In each window, the histogram is modeled by one or two normal distributions, depending on the whether the histogram is unimodal or bimodal. The parameters of the component distributions are estimated. For those windows whose mixture distributions satisfied a bimodality test, minimum error point in the image is then dynamically selected by a linear interpolation procedure.

Sternberg[14] has proposed another dynamic thresholding technique, which is known as the rolling-ball algorithm. In this algorithm, the ball follows the smooth background contours but does not penetrate the object peaks by selecting the ball diameter considerably wider than any peaks formed by the objects in the image thus cannot enter the interior of the peaks. But, the diameter of the ball should be small enough to follow the smooth contours of the changing background intensity. By this process the background image is produced and can be subtracted from the original image to produce the final thresholded image.

2.4 Multi-threshold global thresholding

The global thresholding scheme can be generalized to select multiple threshold values instead of a single threshold value for the whole image. Kohler[15] has proposed a typical thresholding technique with multiple threshold values. In this technique, (n) threshold values are derived from the entire image iteratively. This will divide the feature histogram into (n+1) regions; each is assigned a distinct label. Every pixel in the image is thus assigned a label depending on which region its feature value belongs to.

In general, multi-threshold global methods are better than single global threshold methods. A typical example is the case when feature histogram contains several prominent peaks separated by valleys. In such case, it is not sufficient to select only one global threshold. Usually, it is necessary to select several threshold values that can be determined at the bottoms of the valleys of the gray-level histogram of the image. One typical application for multi-threshold global thresholding method is the cytological image segmentation where three threshold values are computed. These threshold values delimitate background (white), erythrocytes (light gray), cytoplasm (dark gray), and nucleus (dark). [16]

2.5 Thresholding in multi-dimensional space

In many cases, a threshold for classifying each pixel cannot be found with only one feature. However, if two or more features are used, more evidence can be established, leading to a proper segmentation of the image.[17] Ohlander et. al.[18] used a recursive multiple feature thresholding scheme to segment outdoor scenes. Nine features such as the gray-level, color intensity of read, green, and blue, and other features were used in their scheme. In all the nine feature histograms, the one with the best peak was selected for determining the upper threshold value and the lower threshold value. Any pixel within the threshold limits was set to (1) and other pixels were set to (0). In the beginning of this algorithm, the whole image was considered and threshold limits were determined. This splits the image into two or more regions, with each region belong to one of the two groups. These regions were saved and checked one by one for further segmentation. This process is iterated until no sharp peaks can be obtained in all the histograms of the remaining regions. The computation time for this scheme is very expensive although the authors introduced a plan to reduce the computation time.

The key problems with all the preceding thresholding paradigms are the need for multiple passes, expensive computation time, and image storage. Thus all these paradigms are impractical in situations where fast data throughput and practical hardware implementations are essential considerations such as robot vision, military target recognition, and any other real-time applications. These requirements have lead to an interest in variable threshold functions capable of tracking gray level variations in a raster-scan format, thus providing algorithms for real-time implementations. Moreover, the developed algorithms can be integrated on a single monolithic integrated circuit or implemented as an embedded system for real-time applications.

3. VARIABLE THRESHOLD FUNCTIONS

The proposed variable threshold functions represent discrete difference equations often used in digital control system design.[19] These functions are similar to the human learning models derived by many researches years ago.[20, 21] Therefore, one can conclude that learning theory finds a particularly interesting application in thresholding gray-level images for real-time image segmentation[22]. The proposed variable threshold functions in this paper are based on a linear first-order low-pass filter and data hold devices. Four different hold devices are considered, zero-order hold, first-order hold, polygonal hold and slewer hold device. Therefore, four different functions are derived for a single filter transfer function. Many other variable threshold functions can be developed for other filters. However, in this paper, the author limits the derivation of these functions to a simple 1st order low pass filter. Full details about these variable threshold functions are given in the following subsections.

3.1 Variable threshold function based on zero-order hold

The variable threshold function is a discrete difference equation representing the digital transfer function of the thresholding system shown in Fig. 1. The digital system consists of (1) a zero-order hold device to reconstruct the digital data input representing the gray-level values of the image pixels, P[nT], followed by (2) the transfer function of the 1st order linear analog filter, followed by (3) an ideal sampler with a sampling time of (T) seconds, to generate the discrete, variable threshold function, Q[nT], where n=1, 2, ...

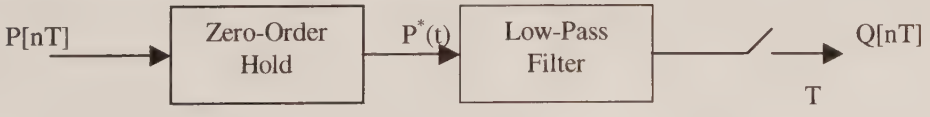


Fig. 1 Zero-Order Hold Thresholding System

The variable threshold function of the zero-order hold thresholding system can be derived using the Z-transform. The transfer function of the zero-order hold device is given as[19]

$$G_{0h}(s) = \frac{1 - e^{-sT}}{s} \quad (1)$$

Where, T is the sampling time in seconds. The transfer function of 1st order linear low-pass filter can be written as

$$G_{LP}(s) = \frac{a}{s + a} \quad (2)$$

Where, a = 1/(RC), R is the resistance of the filter in Ohms, and C is its capacitance in Farads. Therefore, the digital transfer function of the zero-order hold thresholding system can be expressed as

$$D_{0h}(z) = \frac{Q(z)}{P(z)} = Z[G_{0h}(s).G_{LP}(s)] = (1 - z^{-1}).Z[\frac{a}{s(s + a)}] \quad (3)$$

Hence, the corresponding variable threshold function is

$$Q[(n+1)T] = A.Q[nT] + (1 - A).P[nT] \quad (4)$$

Where n = 1, 2, ... and $A = e^{-aT}$. A is a real number with a value between 0 and 1. Equation (4) is the variable threshold function based on a zero-order hold device. The next threshold value is based on the previous threshold value and the current gray-level value of the pixel. The very first threshold value in any trace can be assumed. It is desirable to set the initial threshold value to one-half the gray level value of the first pixel in any trace. A decision can be made in real-time to determine whether the pixel belongs to the object region or the background region based on the current threshold value as follows:

P(I,J) = OBJECT=1	if P(I,J) < Q(I,J)	(5)
P(I,J)=BACKGROUND=0	otherwise	

Equation (4) and equation (5) can be used horizontally and vertically in any image. For horizontal scan, equation (4) is applied on every row individually. In other words, the initial threshold for every row shall be the default value of one-half the gray level value of the first pixel in that row. Similarly, the equations can be used for vertical scan of the image. Either scan shall be sufficient to segment the image.

3.2 Variable threshold function based on first-order hold

Here, the variable threshold function is a discrete difference equation representing the digital transfer function of the thresholding system shown in Fig. 2. The digital system consists of (1) a first-order hold device to reconstruct more accurately the digital data input representing the gray-level values of the image pixels, P[nT], followed by (2) the transfer function of the 1st order linear analog filter, followed by (3) an ideal sampler with a sampling time of (T) seconds, to generate the discrete, variable threshold function, Q[nT], where n=1, 2, ...

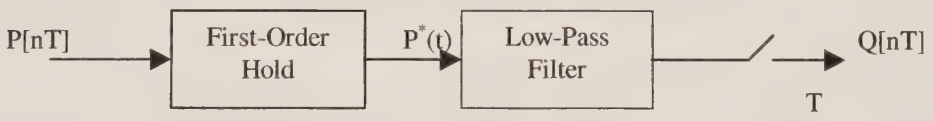


Fig. 2 First-Order Hold Thresholding System

The variable threshold function of the zero-order hold thresholding system can be derived using the Z-transform. The transfer function of the first-order hold device is given as[19]

$$G_{1h}(s) = \left(\frac{1+Ts}{T}\right)\left(\frac{1-e^{-sT}}{s}\right)^2 \tag{6}$$

Therefore, the digital transfer function of the zero-order hold thresholding system can be expressed as

$$D_{1h}(z) = \frac{Q(z)}{P(z)} = Z[G_{1h}(s).G_{LP}(s)] = \frac{z\left(\frac{A}{aT} - A - \frac{1}{aT} + 2\right) + \left(\frac{1}{aT} - \frac{A}{aT} - 1\right)}{z(z-A)} \tag{7}$$

Hence, the corresponding variable threshold function is

$$Q[(n+2)T] = A.Q[(n+1)T] + K_1.P[(n+1)T] + K_2.P[nT] \tag{8}$$

$$K_1 = \frac{1 - A - A.\ln(A) + 2\ln(A)}{\ln(A)} \tag{9}$$

$$K_2 = \frac{A - 1 - \ln(A)}{\ln(A)} \tag{10}$$

Where n = 1, 2, ... and $A = e^{-aT}$. A is a real number with a value between 0 and 1. Equation (8) is the variable threshold function based on a first-order hold device. The next threshold value is based on the previous threshold value and the previous gray-level values of two previous pixels on the same scan. The very first threshold value in any trace can be assumed. It is desirable to set the initial threshold value to one-half the gray level value of the first pixel in any trace. A decision can be made in real-time to determine whether the pixel belongs to the object region or the background region based on the current threshold value as follows:

P(I,J) = OBJECT=1	if P(I,J) < Q(I,J)	(11)
P(I,J)=BACKGROUND=0	otherwise	

Equation (8) and equation (11) can be used horizontally and vertically in any image. For horizontal scan, equation (8) is applied on every row individually. In other words, the initial threshold for every row shall be the default value of one-half the gray level value of the first pixel in that row. Similarly, the equations can be used for vertical scan of the image. Either scan shall be sufficient to segment the image.

3.3 Variable threshold function based on polygonal hold

Here also, the variable threshold function is a discrete difference equation representing the digital transfer function of the thresholding system shown in Fig. 3. The digital system consists of (1) a polygonal hold device to reconstruct the digital data input representing the gray-level values of the image pixels, P[nT], followed by (2) the transfer function of the 1st order linear analog filter, followed by (3) an ideal sampler with a sampling time of (T) seconds, to generate the discrete, variable threshold function, Q[nT], where n=1, 2, ...

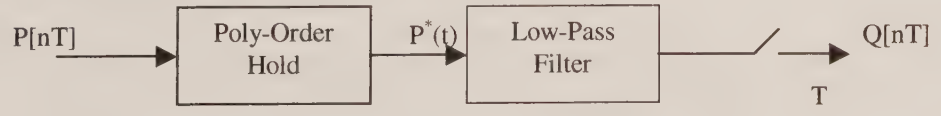


Fig. 3 Polygonal-Order Hold Thresholding System

The variable threshold function of the polygonal hold thresholding system can be derived using the Z-transform. The transfer function of the polygonal hold device is given as[19]

$$G_{ph}(s) = \frac{e^{Ts} + e^{-sT} - 2}{Ts^2} \tag{12}$$

Where, T is the sampling time in seconds. Therefore, the digital transfer function of the zero-order hold thresholding system can be expressed as

$$D_{ph}(z) = \frac{Q(z)}{P(z)} = Z[G_{ph}(s).G_{LP}(s)] = \frac{z(aT - 1 + A) + (1 - A - aAT)}{aT(z - A)} \tag{13}$$

Hence, the corresponding variable threshold function is

$$Q[(n+1)T] = A.Q[nT] + K_3.P[(n+1)T] + K_4.P[nT] \tag{14}$$

$$K_3 = \frac{1 - A + \ln(A)}{\ln(A)} \tag{15}$$

$$K_4 = \frac{A - 1 - A.\ln(A)}{\ln(A)} \tag{16}$$

Where n = 1, 2, ... and $A = e^{-aT}$. A is a real number with a value between 0 and 1. Equation (14) is the variable threshold function based on polygonal hold device. The next threshold value is based on the previous threshold value and the previous two gray-level values of the two preceding pixels in the same scan. The very first threshold value in any trace can be assumed. It is desirable to set the initial threshold value to one-half the gray level value of the first pixel in any trace. A decision can be made in real-time to determine whether the pixel belongs to the object region or the background region based on the current threshold value as follows:

P(I,J) = OBJECT=1	if P(I,J) < Q(I,J)	(17)
P(I,J)=BACKGROUND=0	otherwise	

Equation (14) and equation (17) can be used horizontally and vertically in any image. For horizontal scan, equation (14) is applied on every row individually. In other words, the initial threshold for every row shall be the default value of one-half the gray level value of the first pixel in that row. Similarly, the equations can be used for vertical scan of the image. Either scan shall be sufficient to segment the image.

3.4 Variable threshold function based on slewer hold

Similarly, the variable threshold function is a discrete difference equation representing the digital transfer function of the thresholding system shown in Fig. 4. The digital system consists of (1) a slewer hold device to reconstruct the digital data input representing the gray-level values of the image pixels, P[nT], followed by (2) the transfer function of the 1st order linear analog filter, followed by (3) an ideal sampler with a sampling time of (T) seconds, to generate the discrete, variable threshold function, Q[nT], where n=1, 2, ...

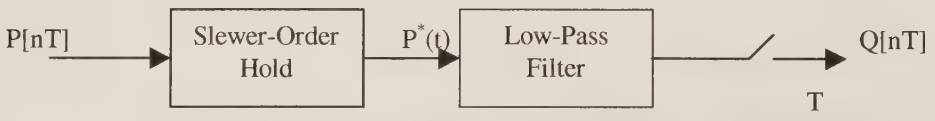


Fig. 4 Slewer-Order Hold Thresholding System

The variable threshold function of the slewer hold thresholding system can be derived using the Z-transform. The transfer function of the slewer hold device is given as[19]

$$G_{sh}(s) = \frac{1}{T}\left(\frac{1-e^{-sT}}{s}\right)^2 \tag{18}$$

Where, T is the sampling time in seconds. Therefore, the digital transfer function of the slewer hold thresholding system can be expressed as

$$D_{sh}(z) = \frac{Q(z)}{P(z)} = Z[G_{sh}(s).G_{LP}(s)] = \frac{z\left(\frac{A}{aT} - \frac{1}{aT} + 1\right) + \left(\frac{1}{aT} - \frac{A}{aT} - A\right)}{z(z-A)} \tag{19}$$

Hence, the corresponding variable threshold function can be defined as

$$Q[(n+2)T] = A.Q[(n+1)T] + K_5.P[(n+1)T] + K_6.P[nT] \tag{20}$$

$$K_5 = \frac{A-1+aT}{aT} \tag{21}$$

$$K_6 = \frac{1-A-aAT}{aT} \tag{22}$$

Where n = 1, 2, ... and $A = e^{-aT}$. A is a real number with a value between 0 and 1. Equation (20) is the variable threshold function based on a slewer hold device. The next threshold value is based on the previous threshold value and the gray-level values of the two previous pixels. The very first threshold value in any trace can be assumed. It is desirable to set the initial threshold value(s) to one-half the gray level value of the first pixel in any trace. A decision can be made in real-time to determine whether the pixel belongs to the object region or the background region based on the current threshold value as follows:

P(I,J) = OBJECT=1	if P(I,J) < Q(I,J)	(23)
P(I,J)=BACKGROUND=0	otherwise	

Equation (20) and equation (23) can be used horizontally and vertically in any image. For horizontal scan, equation (20) is applied on every row individually. In other words, the initial threshold for every row shall be the default value of one-half the gray level value of the first pixel in that row. Similarly, the equations can be used for vertical scan of the image. Either scan shall be sufficient to segment the image.

4. EXPERIMENTS

The developed variable threshold functions have had been tested on a large number of images, with different size, lighting conditions, and characteristics. Many test images were selected, one taken under very low lighting condition, other with abrupt-variation edges, the third with smooth-variation edges, and the fourth with multi-objects. All variable threshold functions produced excellent segmented images on all test images, while the global thresholding technique failed to produce such results. The main difference among the four proposed threshold functions in equations (4), (8), (14), and (20) is the applicable range of the variable A to produce very successful results. For all four threshold functions, it is found that $0.98 \leq A \leq 0.99$ can produce excellent segmentation on all test images. It is also found, that

polygonal-hold threshold function can produce better segmentation over larger range of A values. Table 1 gives full details regarding the range of values of A for successful segmentation for all test images.

	Zero-order hold threshold function	First-order hold threshold function	Polygonal-hold threshold function	Slewer-hold threshold function
Abrupt-variation image	$0.70 \le A \le 0.99$	$0.55 \le A \le 0.99$	$0.64 \le A \le 0.99$	$0.55 \le A \le 0.99$
Smooth-variation image	$0.93 \le A \le 0.99$	$0.92 \le A \le 0.99$	$0.90 \le A \le 0.99$	$0.92 \le A \le 0.99$
Multi-object image	$0.98 \le A \le 0.99$	$0.98 \le A \le 0.99$	$0.97 \le A \le 0.99$	$0.98 \le A \le 0.99$
Poor-lighting condition image	$0.98 \le A \le 0.99$	$0.98 \le A \le 0.99$	$0.98 \le A \le 0.99$	$0.98 \le A \le 0.99$

Table 1 A-values for successful segmentation

Fig. 5a is a 64X64 industrial-part image with 64 gray-level values (0 to 63). The image was taken under natural lighting conditions. Fig. 5b shows the segmented image using the polygonal-hold threshold function. Similar results were obtained with all the other three variable threshold functions.

Fig. 6a represents the same image under low lighting condition. Fig. 6b shows the segmented image using polygonal hold threshold function, and Fig. 6c shows the result of a global threshold function. It is very clear that the segmented image of the proposed approach is superior over global thresholding. Excellent results were also obtained using the polygonal threshold function for images under poor lighting conditions of Fig. 7a, Fig. 8a, and Fig. 9a. The results of the polygonal threshold function are shown in Fig. 7b, Fig. 8b, and Fig. 9b, respectively. On the other hand, bad results were obtained when the conventional global threshold function was applied, as shown in Fig. 7c, Fig. 8c, and Fig. 9c, respectively.

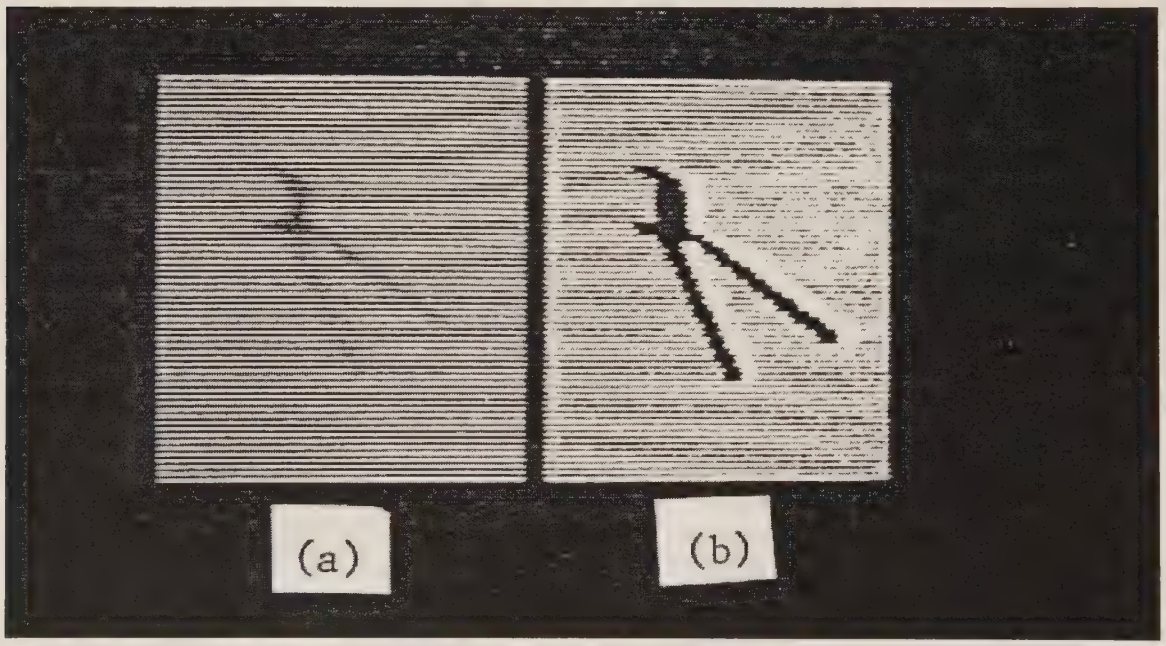


Fig. 5: (a) Industrial-part image under natural lighting conditions, (b) Its segmented image using polygonal hold threshold function.

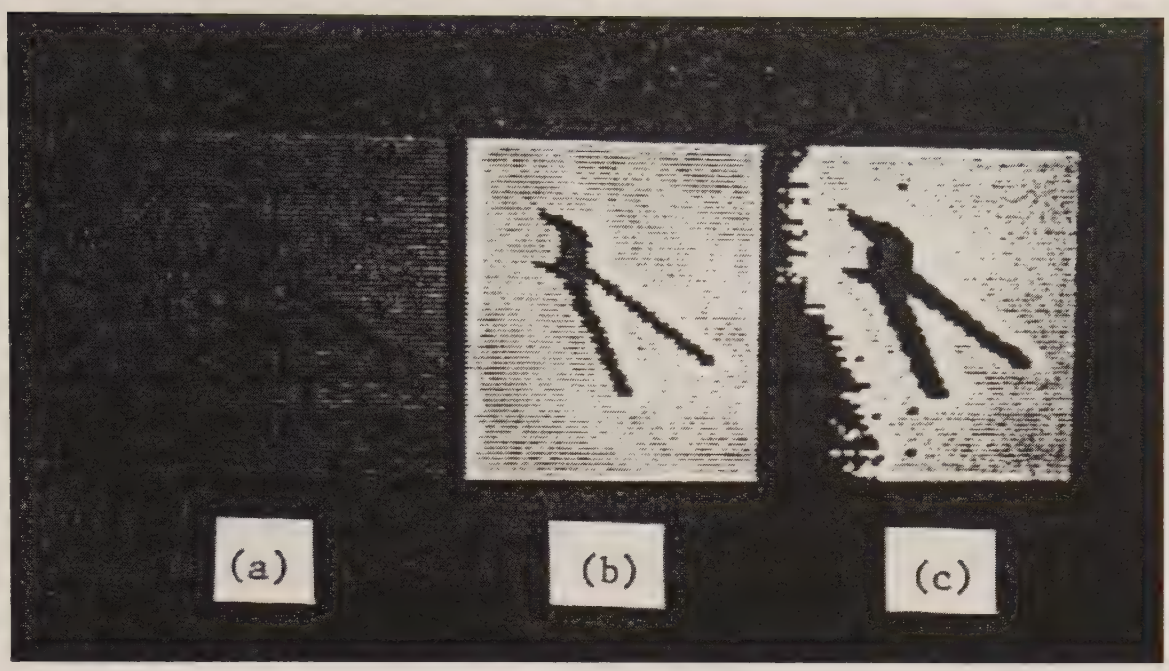


Fig. 6: (a) Industrial-part image under low lighting conditions, (b) Its segmented image using polygonal hold threshold function, (c) Its segmented image using global threshold.

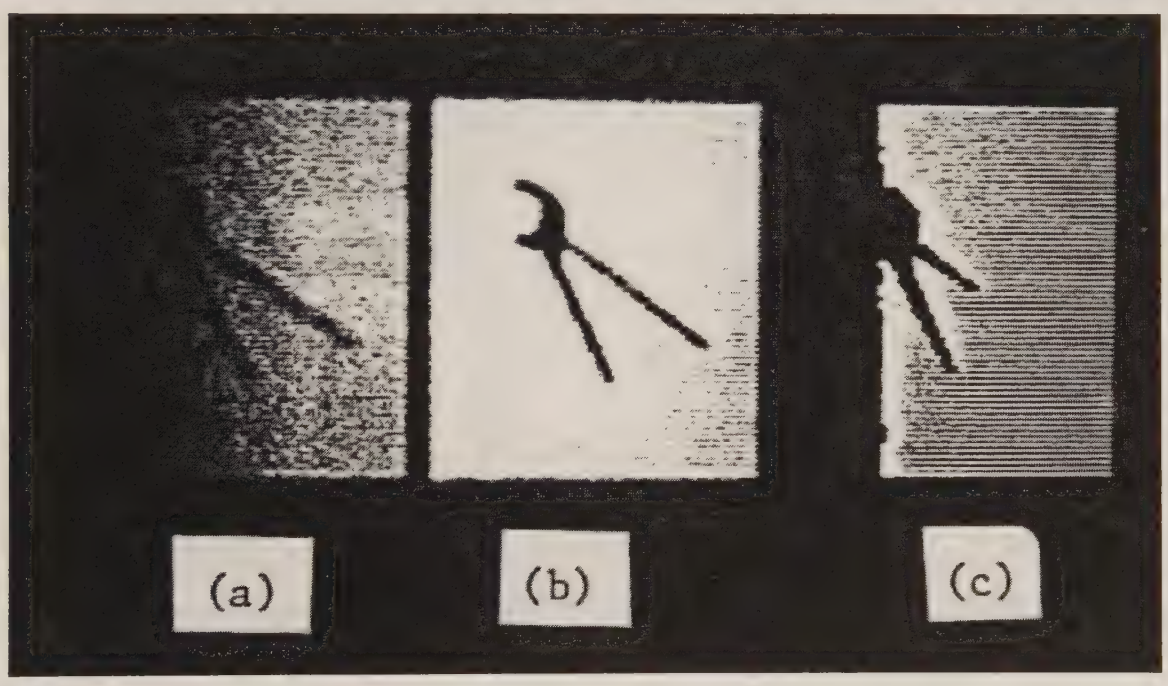


Fig. 7: (a) Industrial-part image under linear lighting function (left-to-right), (b) Its segmented image using polygonal hold threshold function, (c) Its segmented image using global threshold.

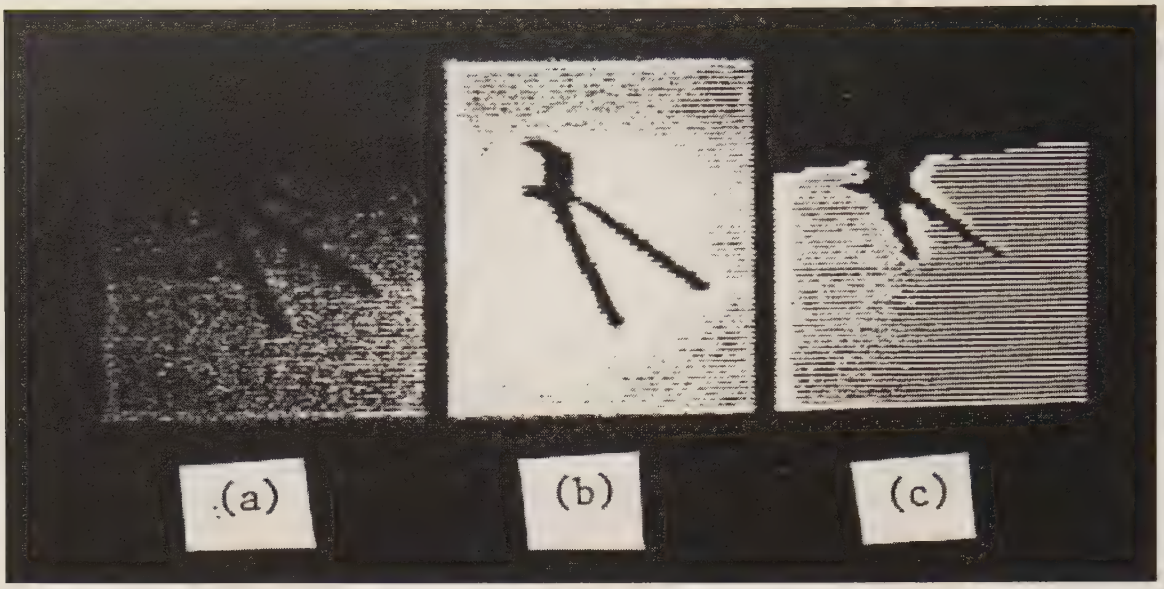


Fig. 8: (a) Industrial-part image under linear lighting function (top-to-down), (b) Its segmented image using polygonal hold threshold function, (c) Its segmented image using global threshold.

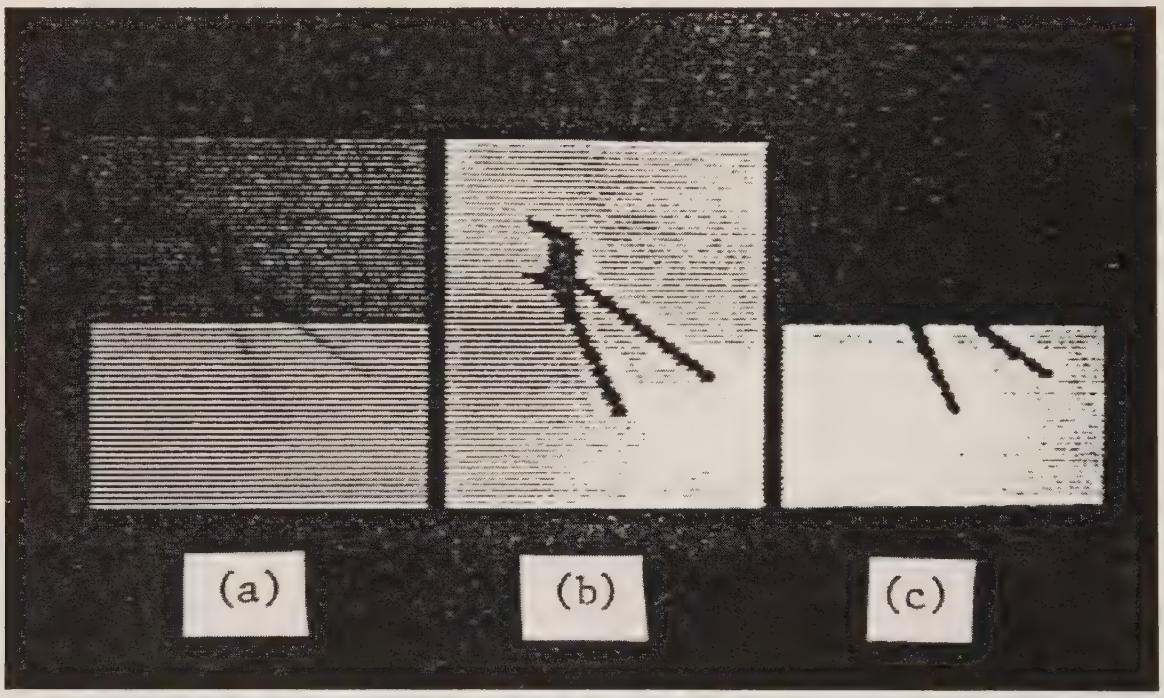


Fig. 9: (a) Industrial-part image under poor lighting conditions, (b) Its segmented image using polygonal hold threshold function, (c) Its segmented image using global threshold.

5. SUMMARY

This paper has developed four different variable threshold functions for real-time image segmentation. All functions have produced successful results and much better than that of their counterparts of global thresholding techniques.

The variable threshold functions have been developed from 1st order low-pass filter and four different hold devices. The hold devices are zero-order hold, first-order hold, polygonal hold, and slewer hold.

The variable threshold functions have been tested on images with different behavior and signal shape. Excellent segmentation has been obtained from all functions but with different range of (A) values. The polygonal-hold threshold function has wider range of (A) values. Moreover, for $0.98 \leq A \leq 0.99$ the entire four variable threshold functions produced very successful results.

This approach is very desirable over all other segmentation techniques because of its hardware implementation. The proposed functions can be implemented on ASIC chip for real-time segmentation and can be incorporated with the digital camera to produce segmented images for practical applications. In addition, the variable threshold function can be implemented on a real-time embedded system.

REFERENCES

1. J.M.S. Prewitt and M.L. Mendelsohn, "The analysis of cell images". Ann. N.Y. Acad. Sci., vol. **128**, pp. 1035-1053, 1966.

2. A. Rosenfeld and A.C. Kak, *Digital Picture Processing*, chapter 8, Academic, New York, 1976.

3. A. Rosenfeld and L.S. Davis, "Iterative histogram modification", IEEE Trans. Syst., Man, Cybern., vol. **SMC-8**, pp. 300-302, 1978.

4. J.S. Weszka and A. Rosenfeld, "Histogram modification for threshold selection", IEEE Trans. Syst., Man, Cybern., vol. **SMC-9**, pp. 38-52, 1979.

5. J.S. Weszka, R.N. Nagel, and A. Rosenfeld, "A threshold selection technique", IEEE Trans. Computer, vol. **C-23**, pp. 1322-1326, 1974.

6. S. Watanabe and the CYBEST Group, "An automated apparatus for cancer prescreening: CYBEST", Comput. Graphics Image Processing, vol. **3**, pp. 350-358, 1974.

7. N. Otsu, "A threshold selection method from gray-scale histograms", IEEE Trans. Syst., Man, Cybern., vol. **SMC-9**, pp. 62-66, 1979.

8. P. Lemkin, "The boundary trace transform: an edge region enhancement transform", Comput. Graphics Image Processing, vol. **9**, pp. 150-165, 1979.

9. A. Rosenfeld and P.D. Torre, "Histogram concavity analysis as an aid in threshold selection", IEEE Trans. Syst., Man, Cybern., vol. **SMC-13**, pp. 231-235, 1983.

10. R.N. Wolfe, "A dynamic thresholding scheme for quantization of scanned images", Proc. Of Automatic Pattern Recognition, National Security Industrial Association, pp. 143-162, 1969.

11. J.R. Ullmann, " Binarization using associative addressing", Pattern Recognition, vol. **6**, pp. 127-135, 1974.

12. C.K. Chow and T. Kaneko, "Boundary detection of radiographic images by a threshold method", Proc. IFIP Congress, vol. **2**, pp. 1530-1535, 1972.

13. G. T. Herman and H.K. Liu, "Dynamic boundary surface detection", Comput. Graphics Image Processing, vol. **7**, pp. 130-138, 1978.

14. S. R. Sternberg, " Biomedical image processing", IEEE Computer, vol. **16**, pp. 22-34, Jan. 1983.

15. R. Kohler, "A segmentation system based on thresholding", Comput. Graphics Image Processing, vol. **15**, pp. 319-338, 1981.

16. C. Garbay, "Image structure representation: a discussion of some segmentation methods in cytology", IEEE Trans. Pattern Anal. Mach. Intell., vol. **PAMI-8**, pp. 140-146, 1986.

17. D. H. Ballard and C. M. Brown, *Computer Vision*, Prentice Hall, New Jersey, pp. 153-155, 1982.

18. K. Ohlander, K. Price, and D. R. Reddy, "Picture segmentation using a recursive region splitting method", ", Comput. Graphics Image Processing, vol. **8**, pp. 313-333, 1978.

19. C. L. Phillips and H. T. Nagle, *Digital Control System Analysis and Design*, 3rd edition, chapter 3, Prentice Hall, New Jersey, 1995.

20. R. R. Bush and F. Mosteller, *Stochastic Models for Learning*, Chapter 5, Wiley, New York, 1965.

21. R. C. Atkinson, G. H. Bower, and E. J. Crothers, *An Introduction to Mathematical Learning Theory*, Chapter 5, Wiley, New York, 1965.

22. H. Hassan and P. Siy, "Thresholding based on learning theory", Proceedings of IEEE, vol. **76**, pp. 1379-1381, 1988.

Dual LCD-based flat panel display

James C. Kirsch* and Brian K. Jones
U.S. Army Research, Development, & Engineering Command, Redstone Arsenal, AL 35898

ABSTRACT

Stereo imagery has been a goal in optics research since the invention of the stereoscope in 1834. While the market has been inundated with displays of various types, sizes, and formats, no general purpose, easy to use, inexpensive method for the display of imagery in stereo has been developed. The benefits of stereo vision are numerous and quickly become apparent when attempting to perform simple tasks without the aid of stereo cues. The proliferation of remotely operated vehicles and indirect view applications has resulted in an increased need to see the operational environment in stereo.

Numerous approaches to the display of stereo imagery have been demonstrated. Stereoscopic displays typically require the user to wear special headgear. Autostereoscopic displays, so named because they do not require the headgear, typically have tight limitations on the position of the viewer's head. Previous papers have described the theoretical underpinnings for new type of stereoscopic displayed based on dual liquid crystal displays. The new display provides a stereo view without temporal or spatial multiplexing. This paper will present the results from experiments to characterize the display components and the resulting changes in the encoding algorithm.

Keywords: Stereoscopic display, LCD, 3D, polarization encoding, flat panel

1. INTRODUCTION

The use of remotely operated vehicles has been rapidly increasing in recent years. Remotely operated, in this case, also refers to situations where an operator rides in the vehicle but is not permitted a direct view of the outside world. There are several scenarios where this may be the case, including situations where the operator's eyes must be protected or where night vision is required. These cases usually require the operator to control the vehicle while viewing the outside world on a monitor or flat panel display. The disadvantage of limiting the viewer to a flat panel display is the loss of many of the depth cues needed to avoid obstacles. Several serious accidents have occurred where the operator misjudged the distance to (or depth of) a ditch in the field of view. One solution to this problem is to provide the operator with stereoscopic vision that restores depth perception. The operating conditions for many applications require the stereo display to be compact and rugged while still providing the necessary depth cues to the driver. Such a system would also be useful for remotely piloted vehicles, remotely operated spacecraft, planetary rovers, deep sea applications, or any other environment dangerous to human operators.

The Weapons Sciences Directorate of the U.S. Army Research, Development, and Engineering Center has recently developed a stereoscopic display based on dual liquid crystal panels such as those found in laptop computers. The basic idea is a polarization encoded display but in a compact package suitable for use in fielded environment. The packaged display should be no larger than standard flat panel displays already in use and only slight thicker. The panels themselves are relatively cheap so the main cost is in providing the stereo (left & right) views of the world to the panels. The advantages of this display approach over other stereoscopic displays are the full frame rate (no temporal multiplexing), full spatial resolution (no spatial multiplexing), and the compact design (no projection optics). The user is required to wear inexpensive polarized glasses. The display can also be used as a regular flat panel display when viewing no stereoscopic imagery or data.

Remotely operated vehicles do not represent the only potential application for 3D vision. Human operators find it easier to detect regions of interest in 3D imagery than in 2D imagery. Applications in this area abound including, but not limited to, medical diagnostic imagery, reconnaissance imagery, and, perhaps, homeland security. The entertainment industry has experimented with different types of 3D imagery for years and appears determined to provide a virtual 3D

* jim.kirsch@us.army.mil; phone (256) 876-8752; fax (256) 955-8283

Optical Pattern Recognition XV, edited by David P. Casasent, Tien-Hsin Chao, Proc. of SPIE Vol. 5437 (SPIE, Bellingham, WA, 2004) · 0277-786X/04/$15 · doi: 10.1117/12.544113

gaming experience in future generations of video games. The nature of the invention described here makes it an ideal candidate for incorporation into laptop computers for high-end applications.

2. DUAL PANEL 3D DISPLAY

Polarization has been used in the past as a tool for separating left and right stereo pairs. Liquid crystal displays (LCDs) are also commonly used for stereoscopic displays. The techniques used previously, however, suffered from complicated optics, reduced resolution, reduced frame rate, bulky packaging, and other drawbacks that made them somewhat undesirable.[1,2] The approach described here attempts to address some of these issues. Two LCDs were stacked together to form a compact panel capable of displaying stereo images without loss of resolution or frame rate. The approach does require the user to wear polarization glasses, similar to ones used in 3D movies.

The proposed architecture is shown in Figure 1. The stereoscopic display was constructed from two Sharp LQ104V1DC31 10.4" LCD panels. A standard backlight provides the necessary illumination for the display. The first panel, closest to the backlight, retains both the input and output polarizers. In essence, this is the standard liquid crystal display. The presence of the polarizers means that irradiance images are visible on the display and, in fact, this display is used to control the total irradiance in the system. A second liquid crystal panel is added to the stack for the stereoscopic display. This panel is identical to the first panel with the exception that both polarizers have been removed. The second panel serves as an array of variable ½ wave retarders that control the polarization orientation of the output light. This panel is used to encode the left and right image data which is then decoded by polarized glasses on the observer.

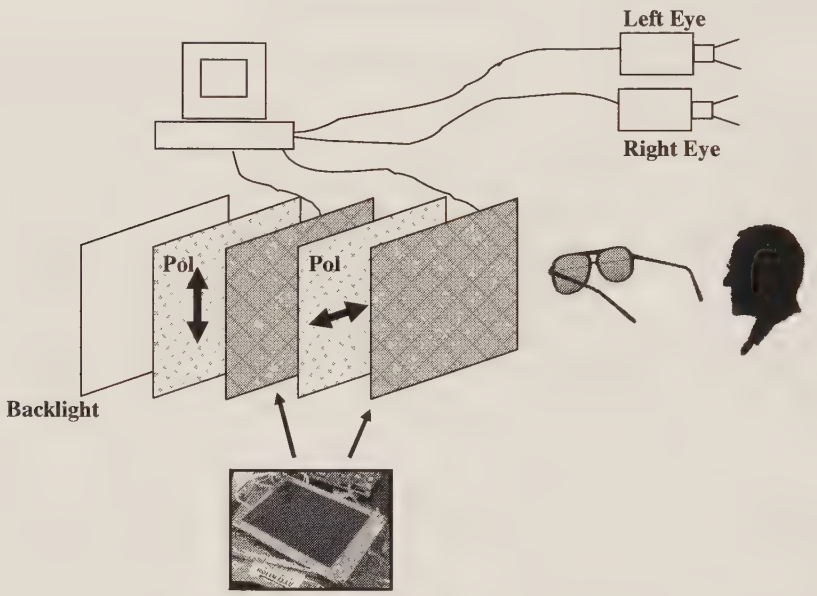


Figure 1. Block diagram of stereoscopic display.

Mathematically, the process is described as follows:

$$I_T = \frac{I_L(x,y) + I_R(x,y)}{2} \tag{1}$$

where $I_L(x,y)$ is the image irradiance distribution captured by the 'Left Eye' camera, $I_R(x,y)$ is the image irradiance distribution captured by the 'Right Eye' camera, and $I_T(x,y)$ is the total irradiance in the resulting integrated image. The

first panel, with both front and back polarizers, is used to encode this intensity image. The images in Figure 2 illustrate the process. Images from the left and right cameras are combined according to Eq. 1 to form the composite image.

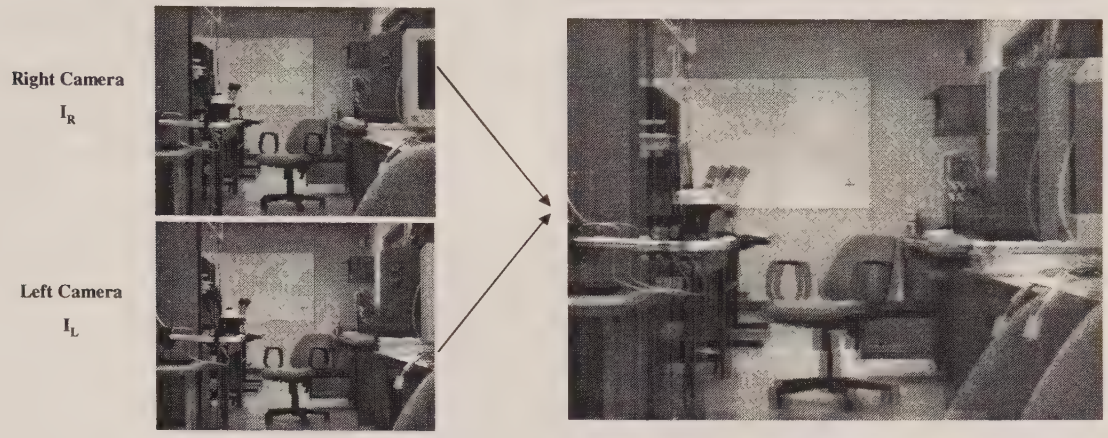


Figure 2. Left and right eye views are combined to form the irradiance image.

Now that the total intensity for the stereo pairs has been encoded, superposition will be used to divide that intensity between the left and right eyes of the observer. Consider the polarization state of the light entering the second LCD panel. All of the light exiting from the first (irradiance) panel and polarizer set has the same polarization as determined by the output polarizer. The second panel has no polarizers but each pixel of the liquid crystal panel rotates the polarization of the incident light by an amount determined by the state of that pixel. The polarization state upon leaving the first panel is illustrated in Figure 3. The polarization state after passing through the second panel at an arbitrary pixel is shown in Figure 4. The amplitude of the light leaving the first panel determines the length of the electric field vector (E_T). The orientation of that vector is determined by the amount of rotation imparted by the second panel. Superposition can be used to write the orientation of that vector in terms of orthogonal polarizations referred to as $E_L(x,y)$ and $E_R(x,y)$. The polarizing glasses worn by the observer will now preferentially pass these two components to the appropriate eye. Therefore, how the light is divided between the two eyes is determined by the orientation of the electric field vector (E_T) leaving the second panel. Changing the orientation of the electric field vector changes the amount of light reaching each eye. In this manner, the light leaving the second panel can be directed to the appropriate eye for the left and right stereo pairs.

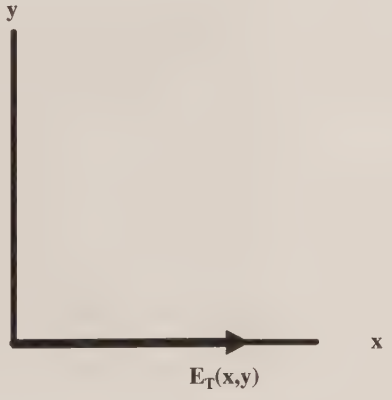


Figure 3. Polarization orientation of light leaving irradiance panel and polarizer.

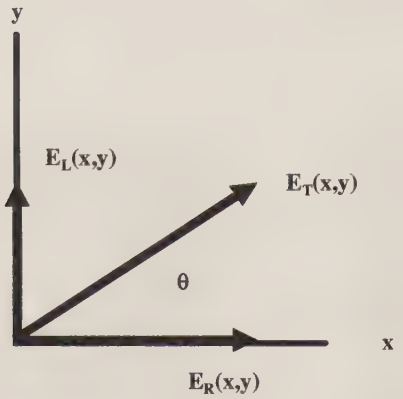


Figure 4. Polarization orientation of light leaving an arbitrary pixel of the polarization panel.

From Malus's Law, the irradiance of the light passing through an analyzer can be written as:

$$I_R(x,y) = I_T(x,y)\cos^2\theta = \left[I_L(x,y) + I_R(x,y)\right]\cos^2\theta \tag{2}$$

$$\cos^2\theta = I_R(x,y) / \left[I_L(x,y) + I_R(x,y)\right] \tag{3}$$

where I_L is the irradiance in the left image, I_R is the irradiance in the right image, and θ is the polarization angle from Figure 4. This illustration assumes the transmission axis of the analyzers in front of the left and right eyes are oriented along the x-axis and y-axis, respectively. Figure 5 is an example 'polarization image' derived using Eq. 3. Comparing the image with the diagram in Figure 4, yields confidence that the algorithm is working as expected. Bright areas of the image correspond to most of the light going to one eye, while darker areas correspond to most of the light directed to the other eye. Regions of medium gray are areas where the light is approximately divided evenly between the two eyes. As an example, compare the position of the chair arm in the top and bottom images. It is easy to see that the chair arm in the top image appears farther to the left than the arm in the bottom image. In the composite polarization image, the arm appears twice - darker on the left and brighter on the right. This corresponds to directing the light from the arm to different eyes. The resulting disparity gives rise to the perception of depth.

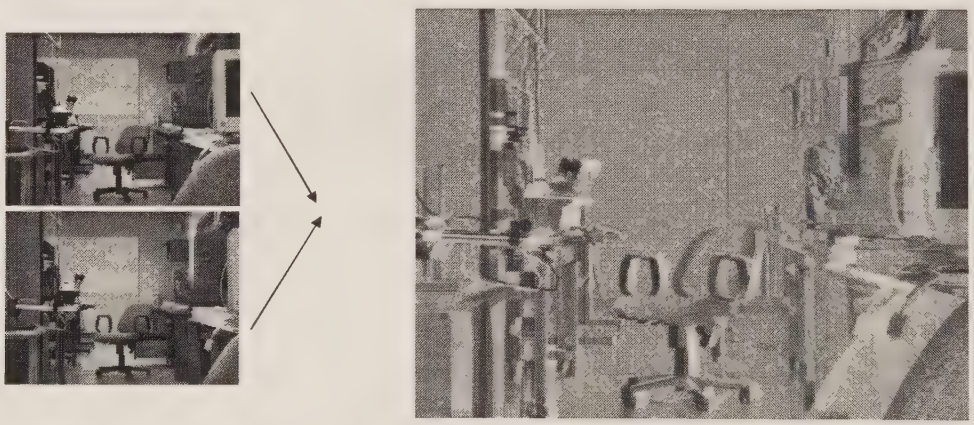


Figure 5. Left and right eye views combined to generate 'polarization' image. White areas correspond to all light going to one eye while dark areas correspond to all light going to the other eye.

4. COMPONENT EVALUATION

The panel chosen for this initial work, the Sharp LQ104V1DC31, is shown in Figure 6. This is a 10.4" diagonal panel with 640x480x3 pixels, where the pixels are arranged in an RGB fashion to produce a color image. The initial experiments were performed with grayscale images; however, the technique can easily be applied to produce a color image. The panels can be individually driven with either a standard NTSC video signal or a digital 640x480 computer video signal through a Digital View ACL-1024 video controller. The performance of the individual panels will determine the overall performance of the 3D display as well as dictate how the algorithm must be modified to maximize performance.

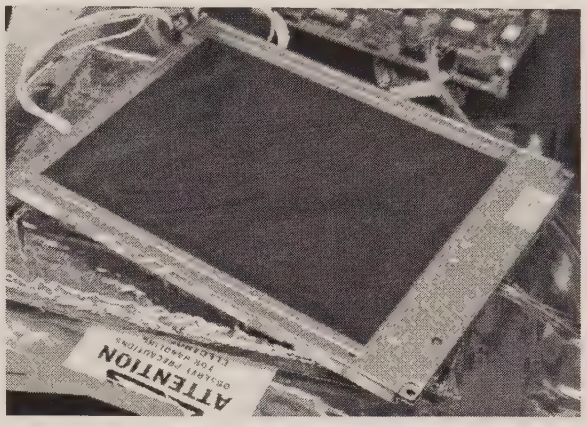


Figure 6. Photograph of the Sharp LQ104V1DC31 liquid crystal panel.

The algorithm described above makes two key assumptions about the modulation characteristics of the liquid crystal panels. First, it assumes that the panels are capable of providing a variable 0 to 90 degree change in the orientation of the polarization vector. Second, it assumes that the modulation curve is linear. Neither of these assumptions is likely to be correct in an off-the-shelf device. Initial attempts to generate the stereoscopic images with the Sharp panels and the algorithm discussed were sufficient to prove the concept, but also indicated that a better understanding of the modulation characteristics was required. The stereoscopic images were discernible however, significant ghosting was also present. Ghosting refers to seeing left eye image data with the right eye and vice versa.

A Stokes vector imaging polarimeter was constructed to analyze the modulation characteristics of the liquid crystal panels. A diagram of the polarimeter is shown in Figure 7. This particular configuration is frequently referred to as a rotating waveplate polarimeter. The polarimeter consists of a rotating ¼ waveplate, an fixed linear polarizer, and an imaging detector. The rotating waveplate is placed at an intermediate image plane to minimize wedge effects. The calibration procedure and methods for analyzing the data to determine unknown polarization states are discussed in the literature.[3]

The Stokes vector polarimeter is designed to determine the polarization state of the unknown input light. In this particular case, the unknown polarization state is a result of the liquid crystal panel. Linearly polarized light at a known orientation is incident on the liquid crystal panel with no polarizers. Relay lenses between the liquid crystal panel and the rotating waveplate are not shown. The Stokes vector of the light after passing through the panel was determined as a function of pixel grayscale as well as brightness and contrast. Lenses in the system magnify the image of the liquid crystal panel such that individual pixels are visible on the detector. It is important to note that the color filter in the panel will affect the results. At a minimum, the retardance of the waveplate is a function of wavelength. The individual pixels were imaged onto the detector so that data could be collected for red, green, and blue pixels in order to account for wavelength in the characterization. The input polarized light was provided by a Labsphere USS-600 uniform illumination source. The uniform source provides white light with a uniform luminance over a 2" diameter aperture. The light is then polarized by a Melles Griot dichroic sheet polarizer like the 03 FPG 007.

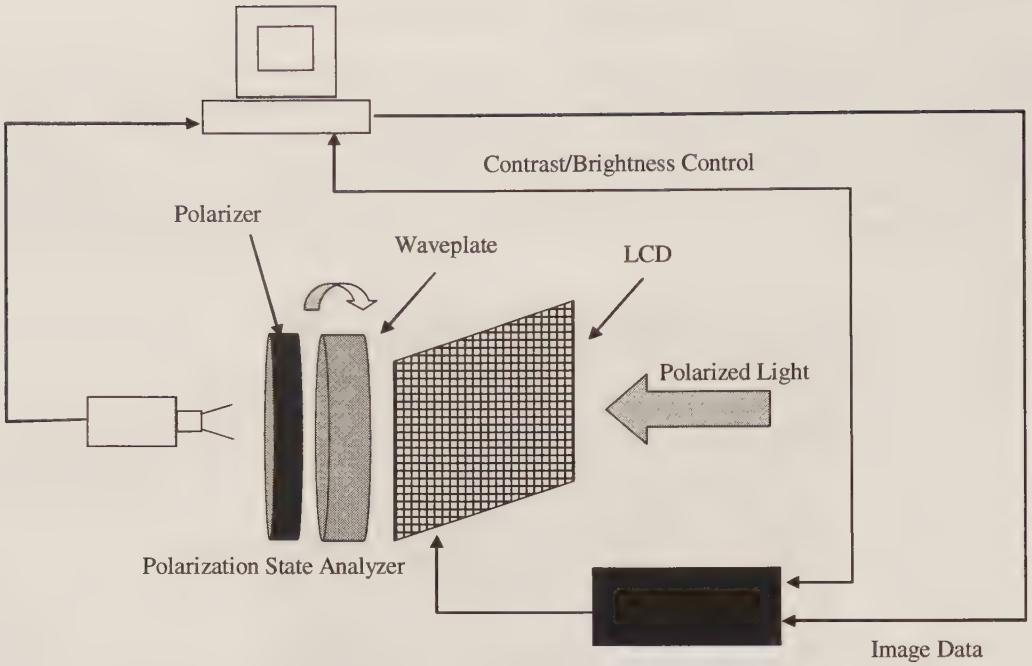


Figure 7. Diagram of the Stokes vector polarimeter used to characterize the liquid crystal panels.

The first step in determining the polarization modulation characteristics of the LCD was to determine the output polarization state from the intensity panel. The intensity panel was placed in the polarimeter and the detector captured sixteen intensity images as the waveplate was rotated through 360 degrees. An analysis of the data was performed to determine the Stokes vector for the light leaving the intensity panel. This polarization state was used as the starting point for calibrating the polarization (top) panel of the stereoscopic display.

The polarization panel was placed into the polarimeter with the input polarization state aligned with the output state from the intensity panel. The Stokes vector of the light leaving the polarization panel as a function of input pixel level was determined for a variety of brightness and contrast settings on the LCD. Several things were immediately apparent from this data. First, the achievable polarization rotation was approximately 70 degrees rather than the desired 90 degrees. Second, the circular polarization component of the light leaving the panel increased significantly with pixel value. Both of these conditions would lead to significant noise in the 3D image using the theoretical equations described above. Noise, in this case, is defined as unwanted light reaching the eye making it difficult to achieve an 'off' state in a particular eye.

The data collected using the manufacturer's original polarizer settings is shown here. Figure 8 is a plot of the polarization ellipse as a function of grayscale. The light leaving the panel for the lower grayscale values is relatively linear. As the pixel value increases, the light becomes more elliptical and then begins to return to a linear state. Even at a grayscale value of 255 however, the light is still fairly elliptical. It is also evident that the orientation of the polarization vector does not rotate the full 90 degrees as expected. Figure 9 is a plot of the irradiance at each eye as a function of grayscale. The graph was generated by multiplying the Stokes vector at each pixel grayscale by the Mueller matrix of a polarizer at the left and right eye orientation. The contrast between the left and right eye at a grayscale of 0 is very high. The contrast at the other end of the grayscale range is much lower. The net result is that one eye will see a very high contrast image while the other eye will lower contrast and significant ghosting. The ideal result would be for the contrast ratio to be very high and balanced on both ends of the curve.

The plots in Figure 8 and Figure 9 led to two conclusions. First, with less than a 90 degree change in the orientation of the polarization vector, orthogonal polarizers on the eyes are not the optimal choice. Contrast is primarily of function of

how dark the "black" or "off" state is in the image. It is readily apparent that crossed polarizers were chosen by the manufacturer because the most linear state, and hence the darkest "off", occurs when the maximum voltage is applied to the panel. For the stereoscopic display, good "off" states are required on both ends of the grayscale curve. The optimal orientation for the eye polarizers is, then, orthogonal to the major axis of the polarization ellipse for the 0 and 255 graylevels. One eye polarizer is chosen to be orthogonal to the major axis of the ellipse at graylevel 0 and the other eye polarizer chosen to be orthogonal to the major axis of the ellipse at graylevel 255.

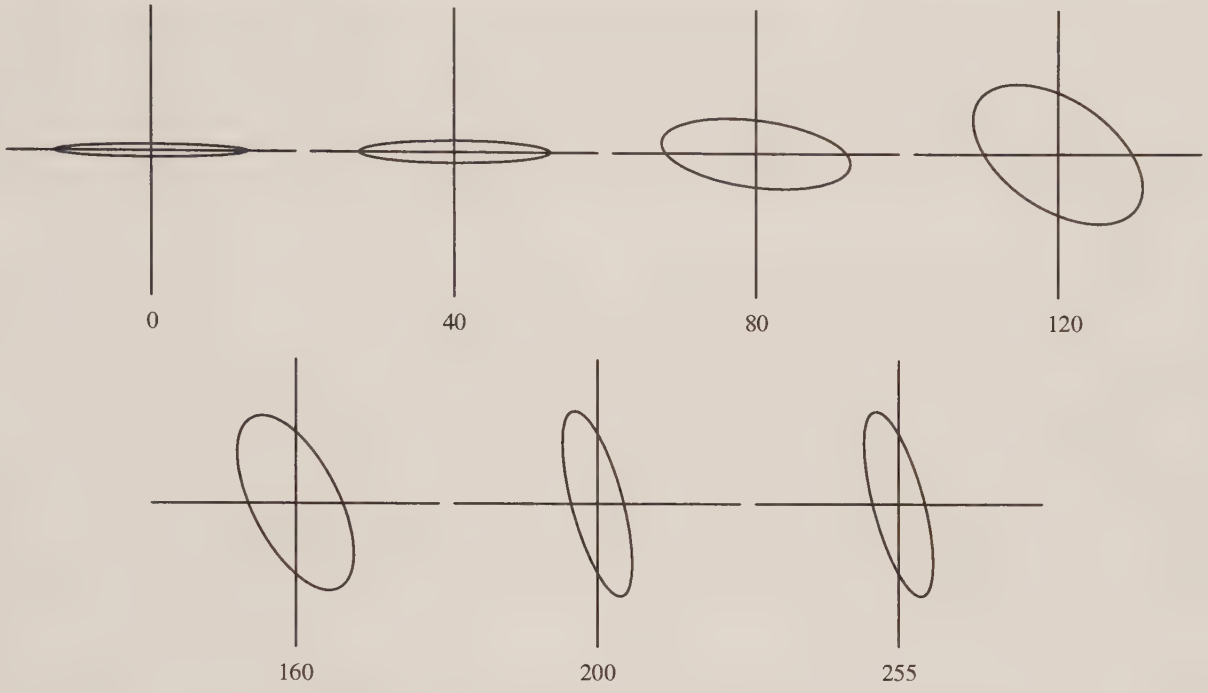


Figure 8. Polarization ellipses with manufacturer's original polarizer orientations.

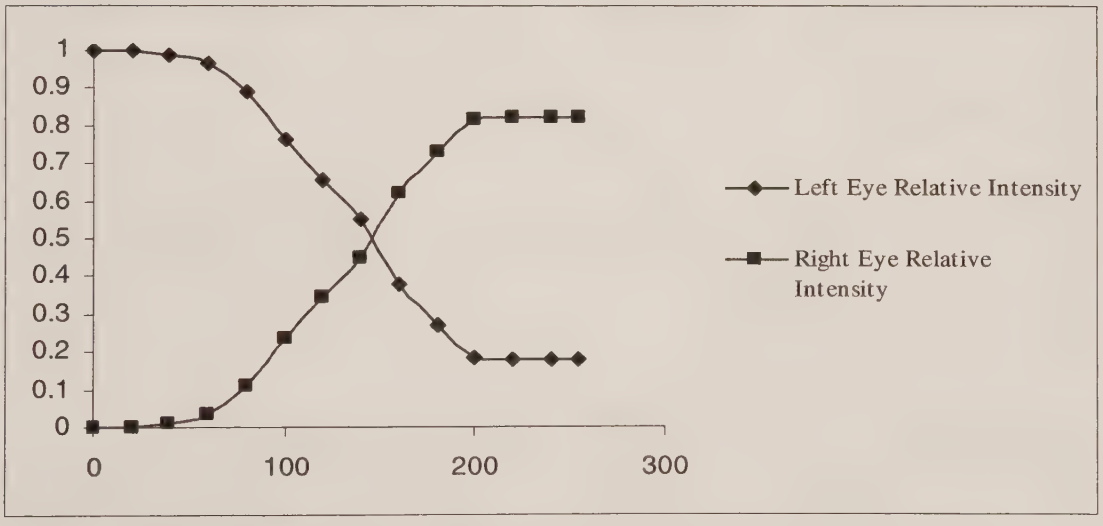


Figure 9. Relative irradiance reach the left and right eyes as a function of pixel value.

The second conclusion was that the amount of ellipticity in the polarization vector needed to be minimized for best contrast. It was theorized that the original manufacturer's polarizer orientation was not aligned with the liquid crystal director. The Stokes vector data was collected for a variety of input polarizer orientations. Figure 10 is a plot of the polarization ellipses for the optimal input polarization orientation. The ellipticity of the polarization has been minimized, particularly at the upper end of the grayscale range. The plot in Figure 11is the relative irradiance reaching each eye for this input polarization setting and with the optimized eye polarizer settings. The contrast is high on both ends of the grayscale.

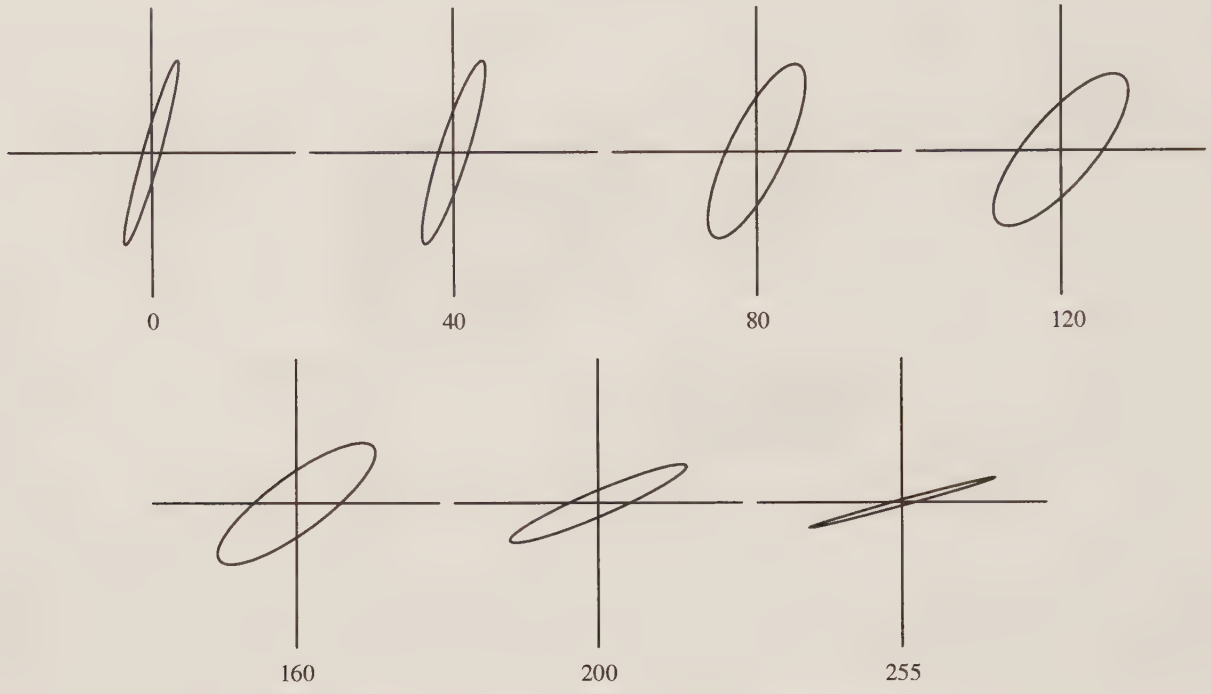


Figure 10. Polarization ellipses for the optimized input polarization orientation.

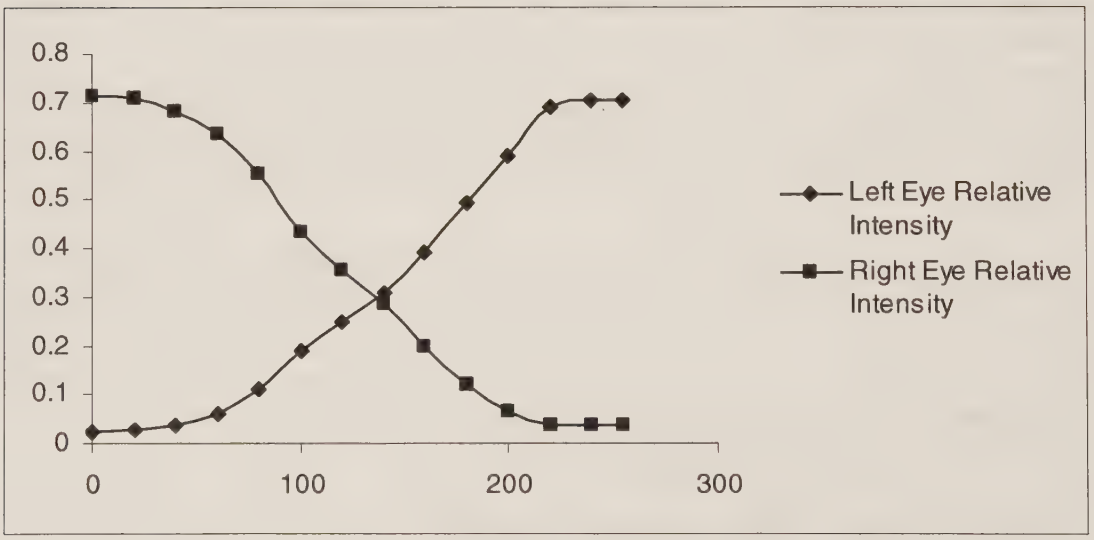


Figure 11. Relative intensity reaching left and right eyes as a function of pixel value.

The effects of brightness and contrast were also studied. Referring to Figure 11, changes in the brightness settings cause the graph to slide left and right along the x-axis. Brightness acts as a bias that effectively sets were the pixels begin to "switch on". Contrast changes the slope of the line. Brightness and contrast settings have been optimized in Figure 11 to minimize the flat regions at the top and bottom of the curve.

5.0 ALGORITHM MODIFICATIONS

The curve in Figure 11 is quite clearly not the ideal curve envisioned by the original algorithm. The nonlinearities in the curve can easily be taken into account by using the curve to generate look-up tables (LUTs) for the polarization panel. The curve shown in Figure 11 is for the green pixels. Similar, but different, curves were also generated for the red and blue pixels. All of the curves have the same general shape but there are differences in slope. LUTs were generated for all three color planes (red, green, and blue). The irradiance panel curve was measured by simply inserting a power meter after the panel and measuring the relative irradiance as a function of grayscale. This curve is also used to generate a LUT for the irradiance panel.

One more factor related to the less than ideal modulation characteristics of the panels must be taken into account. The most demanding situation, in terms of available light, occurs when both the left and right images have a fully on (pixel value = 255, normalized to 1.0) pixel. The same amount of light needs to reach both the left and right eyes, which occurs at the crossing point in Figure 11. The relative intensity reaching both eyes at that point, however, is only 29% of the total light available according to Figure 11. Now consider the case where the left image pixel value is 255 (normalized to 1.0) and the right image pixel value is 0. From Eq. 1, the relative irradiance value would be 0.5. From Figure 11, the polarization panel relative irradiance value would be 0.71. The relative irradiance reaching the left eye is then 0.5*0.71 = 0.36. Now, in both cases, the left image value is the same and the same amount of light should reach the left eye from both pixels. The analysis shows, however, that more light is reaching the left eye when only the left image pixel is on (0.36) than when both the left and right pixels are on (0.29). The irradiance panel value must be modified so that a 255 in the left image yields the same amount of light in the left eye irrespective of the right image pixel value. The same holds true when considering the right eye.

The polarization panel is correctly dividing the light between the left and right eyes. The low transmission values result from the fact that the polarizer on each eye is oriented to provide the best "off" or black state rather than the maximum transmission. If a full 90 degree rotation was present in the panel, most of the problem would disappear. The low transmission value is unavoidable given the less than 90 degree twist. A scaling factor must therefore be applied when determining the appropriate irradiance panel setting.

The total transmitted light reaching both eyes in the ideal case is given by:

$$I_{ideal} = \frac{I_L + I_R}{2*255} \tag{4}$$

where I_{ideal} is the total light reaching the eyes scaled from 0 to 1. The actual light reaching the two eyes (neglecting, for the moment, the nonlinearity in the irradiance panel) is given by:

$$I_{actual} = \frac{I_L + I_R}{2*255} * (L+R) \tag{5}$$

where I_{actual} is the actual total irradiance reaching both eyes, and L and R are the left and right eye transmission values from Figure 11. The second term in this equation takes into account the lower than ideal transmission values in the polarization panel/polarized glasses combination. The changing ellipticity in the polarization from the panel, as well as the non-orthogonal polarizers on the eyes, means that the quantity $L+R$ is no longer a constant. The input irradiance must be scaled to account for these factors. The maximum value of the actual total irradiance reaching both eyes can be no greater than the worst case (both pixels are 255) described above. The scaling factor is, therefore:

$$ScaleFactor = \frac{2 * T_{cross}}{(L+R)} \tag{6}$$

where T_{cross} is the transmission at the crossing point in Figure 11. The procedure, then, is to calculate the irradiance value using Eq. 1 and then scale it using Eq. 6.

6.0 STEREOSCOPIC DISPLAY

A photograph of the stereoscopic display is shown in Figure 12. The two panels have been sandwiched together with the polarizers on the irradiance panel aligned in the optimal configuration determined above. The algorithm has been encoded in look-up-tables to improve the performance of the software. A stand-alone hardware implementation of the electronics has been designed and is in fabrication. The stand-alone version should allow real time (30 Hz) frame rates with stereo cameras providing the input imagery.

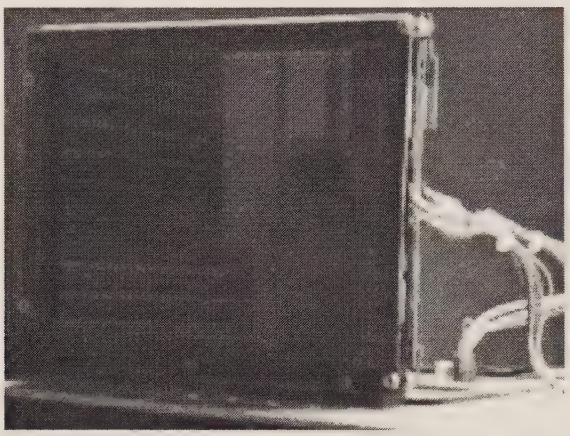


Figure 12. Photograph of the assembled dual flat panel display.

A sample stereo pair is shown in Figure 13. The images are printed with the left image on right and the right image on the left. This is sometime referred to as the cross-eye arrangement. Viewers can see the stereo effect by allowing their eyes to cross in front of the page. The resulting virtual image that appears between the two real images should have depth. The images in Figure 13 were processed according to the algorithm described above and sent to the appropriate panels of the display. The images shown in Figure 14 where taken by viewing the stereoscopic display through the appropriate left eye and right eye polarizers. The images are again printed in the cross-eye format to allow the reader to see that the depth information has been retained.

Figure 13. Original stereo pairs in cross-eye arrangement. Images courtesy of Andrew Woods, The Centre for Marine Science & Technology at Curtin University.

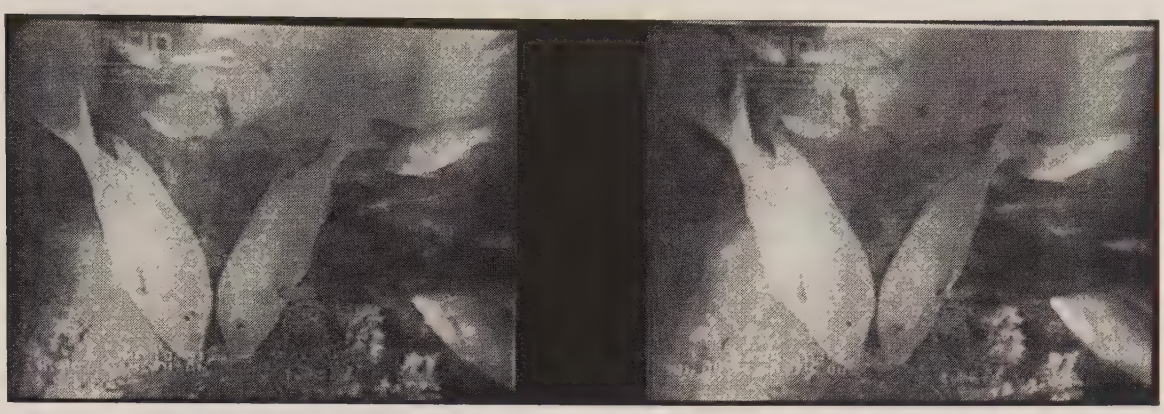


Figure 14. Photograph of fish as displayed on LCD stack and seen through the left and right eye polarizers. The left eye image is on the right so that tthe stereo effect can be viewed using the cross-eye technique.

The images in Figure 14 have a discernible fringe pattern across them. The fringe pattern is a result of crosstalk between pixels in the stacked arrangement. The separation between the two panels allows light from a pixel in the bottom panel to pass through an adjacent pixel in the top panel. The effect is magnified by the color filters on the panels in the sense that light passing through a blue pixel in the bottom panel might pass through an adjacent green or red pixel in the top panel.

7.0 CONCLUSION

The dual LCD flat panel display concept presented here has been validated with a working prototype. The results from the characterizations described have led to a greatly improved stereo display over the original algorithm. Procedures for determining the optimal polarizer settings for the polarization panel and the glasses as well as the required algorithm modifications have been discussed. The display offers the advantages of full resolution and full frame rate while still maintaining a rugged and compact design.

The near term efforts will be focused on evaluating the display in some real world applications. Human-machine interface issues, particularly those related to 3D perception, will be key factors. Future efforts will examine ways to eliminate or reduce the fringing effects and will also evaluate other liquid crystal panels for improved performance.

REFERENCES

1. B. Lane, "Stereoscopic Displays," in *Processing and Display of Three-Dimensional Data*, J.J. Pearson, ed., Proc. of the SPIE 367, 22-23 (1982).
2. J. Chen, K.H. Kim, N.D. Kim, J. Souk, P.J. Bos, and S.T. Shin, "Simple Multimode Stereoscopic Liquid Crystal Display", *Japanese Journal of Applied Physics*, **36**, 1685-1688 (1997).
3. D. B. Chenault, *Infrared Spectropolarimetry*, 21-30, University of Alabama in Huntsville (1992).

Improved Target Detection Using Polarization-Enhanced Fringe-adjusted Joint-Transform Correlation

Aed M. El-Saba, M. S. Alam, E. H. Horache, and S. Regula

Department of Electrical and Computer Engineering
University of South Alabama
Mobile, AL 36688

Abstract

We introduce novel optoelectronic target detection technique using polarization enhancement. Images correspond to elements of Stokes vector imagery are introduced as input scenes to the fringe-adjusted joint-transform correlator. Our results show excellent improvement in performance parameters. Both computer simulation and experimental results are presented in support of the proposed technique.

1. Introduction

There has been considerable interest in target detection and discrimination using imaging polarimetry [1-6]. Since polarization parameters, are generally more sensitive to the nature of scattering surface than the total intensity, it would seem that imaging polarimetry techniques offer a distinct possibility of yielding images with higher visual contrast than traditional total intensity imaging techniques. This is an important factor in several applications of interest.

Most previous investigations [1-6] considered active techniques using IR wavelengths, with little attention paid to passive imaging using visible wavelengths. In this paper we use the concept of visible passive imaging polarimetry to improve target detection and discrimination.

Since most field deployment applications require target detection and discrimination in real time the concept of joint-transform correlation (JTC) [9,10] has been found to be inherently suitable for such applications. Among the various types of JTCs, the fringe-adjusted JTC (FJTC) [11,12] has been found to yield significantly better correlation output compared to alternate JTCs.

We simultaneously utilize the concepts of FJTC and polarization-enhancement to significantly improve correlation performance of target detection and discrimination.

Section 2 contains brief discussion of imaging polarimetry using the formulism of Stokes vector imagery (SVI) and definition of its elements. Section 3 introduces the basic architecture of the FJTC used in our experiment. Section 4 contains simulation and experimental results. Correlation parameters are introduced in this section followed by analysis and comparison of results obtained. Section 5 contains a conclusion and future investigation.

Optical Pattern Recognition XV, edited by David P. Casasent, Tien-Hsin Chao, Proc. of SPIE Vol. 5437 (SPIE, Bellingham, WA, 2004) · 0277-786X/04/$15 · doi: 10.1117/12.543237

2. Imaging Polarimetry and Stokes Vector Imagery

The objective of imaging polarimetry is to determine the polarization content associated with each pixel in the scene. The most general and widely used formulism of imaging polarimetry is the SVI [7,8]. In the SVI representation the polarization properties of the electromagnetic radiation can be completely described by the Stokes vector imagery $\mathbf{S_I}$ that consists of four images known as the Stokes images:

$$\mathbf{S_I} = [S_0, S_1, S_2, S_3]^T. \quad (1)$$

In Eq. (1) T designates the transpose of $\mathbf{S_I}$. S_0 image is an intensity image and it does not carry any polarization information. Generally, an image based on S_0 is equivalent to one obtained by conventional polarization-blind imaging systems that measure intensity only. S_1, S_2, and S_3, images, respectively describes linear polarization, orientation, and circular polarization content of light in the scene. SVI can be determined using a time-sequential intensity measurement approach [1,6,8] as in the following arrangement:

$$S_0 = I(0^0) + I(90^0). \quad (2)$$
$$S_1 = I(0^0) - I(90^0). \quad (3)$$
$$S_2 = I(45^0) + I(135^0). \quad (4)$$
$$S_3 = I(\text{Rcirc}) + I(\text{Lcirc}), \quad (5)$$

where $I(0^0)$, $I(90^0)$, $I(45^0)$, $I(135^0)$, I(Rcirc), and I(Lcirc) are respectively the measured intensity at 0^0, 90^0, 45^0, 135^0, right-circular, and left-circular polarizations.

3. Fringe-adjusted JTC and Correlation Performance Parameters

An architecture for optical implementation of the proposed FJTC is shown in Fig. 1. A laser illuminates the input plane SLM, which simultaneously displays the unknown input image of a target and the reference image to form the joint input image. The input joint image is Fourier transformed by lens L1 and the resulting joint power spectrum (JPS) is detected by the CCD camera. The JPS is multiplied by a real valued filter, called the fringe-adjusted filter (FAF), to yield the fringe-adjusted JPS. This step is achieved by displaying the JPS and the FAF side by side using two separate SLMs and illuminates them simultaneously with the same laser, utilizing a beam splitter and a mirror combination, as shown in Fig. 1. An inverse Fourier transform by lens L2 yields the correlation output.

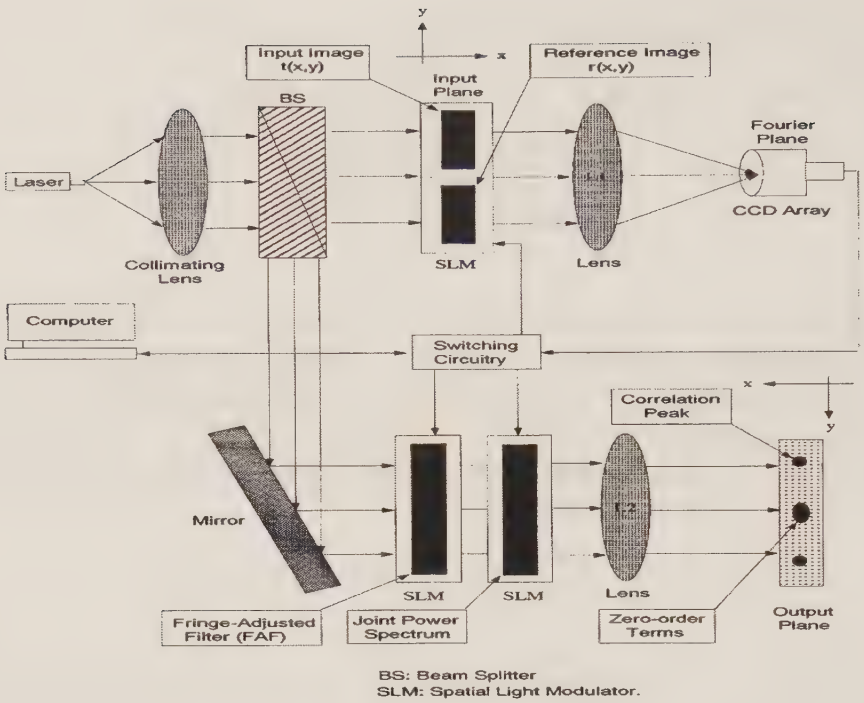


Fig. 1 Arrangement for FJTC

To measure correlation performance, the following three parameters are used:

1- The peak value of correlation energy (P).
2- Peak-to-side lobe ratio (PSR), defined as

$$PSR = (P-m)/\sigma, \tag{5}$$

where m is the mean value of a 3x3 pixels area around the peak (excluding the peak) and σ is the standard deviation of the same. In general the higher the value of PSR the sharper is P.

3- Peak-to-correlation energy (PCE), defined as

$$PCE = P/E, \tag{6}$$

where E is the total energy contained in the correlation plane. PCE provides a measure of peak sharpness and correlation efficiency.

4. Computer Simulation and Experimental Results

Fig. 2(a) shows an input joint image of a black color military truck placed on a dark gray background platform. The selection of this combination of target/background is to provide a minimum contrast between the target and background. Fig. 2(a) is captured by traditional intensity imaging system that is equivalent to S_0 image in the SVI, as defined in Eq. (2).

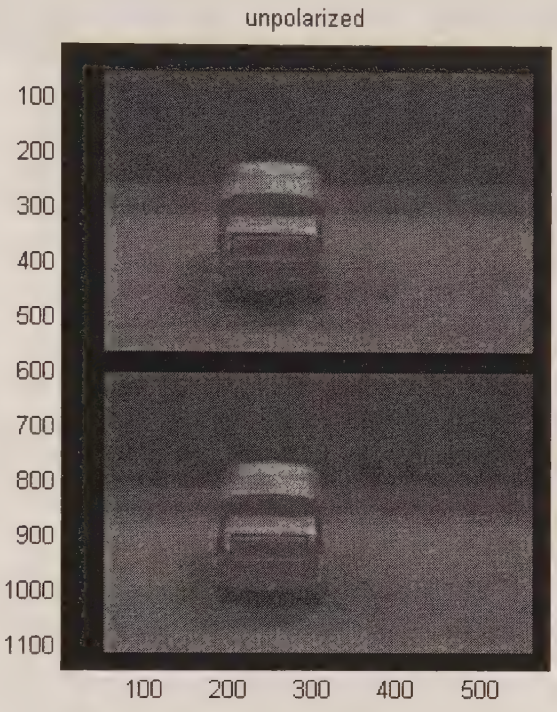


Fig. 2(a): Input joint image of black color military truck on a dark gray background.

This image is applied to the FJTC and the correlation output is plotted in Fig. 2(b). Fig. 2(c) is a two-dimensional plot of Fig. 2(b). We note that in our simulation results Fourier plane subtraction is used to eliminate the dc term.

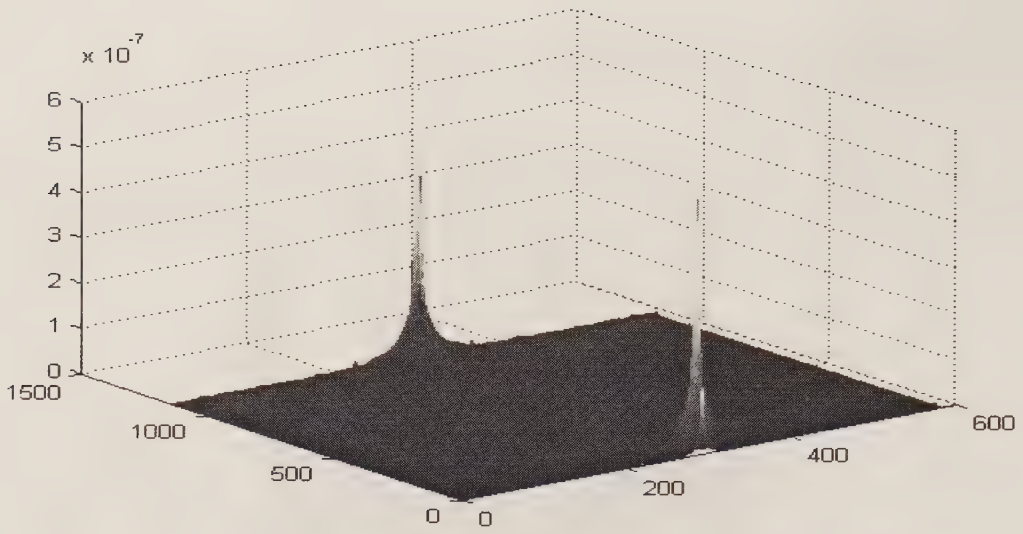


Fig. 2(b): Correlation output of FJTC

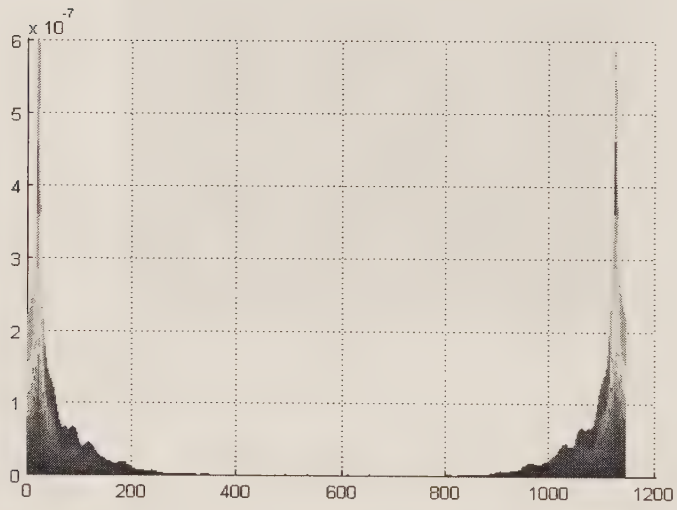


Fig. 2(c): 2-dimenstional view of Fig. 2(b).

The correlation performance parameters for this case are: P ≈ $6x10^{-7}$, PSR ≈ 9.65, and PCE ≈ 0.4%.

Fig. 3 shows the experimental results correspond to this case. Figure 3 indicates (very hard to see) a very faint pair of target peaks in the first and third quadrant (counterclockwise direction) of the correlation output plane. The two peaks are completely overshadowed by the dc terms. Note in particular that PCE ≈ 0.4% indicating a negligible peak. This will result in a high probability of target misses and false alarm rate.

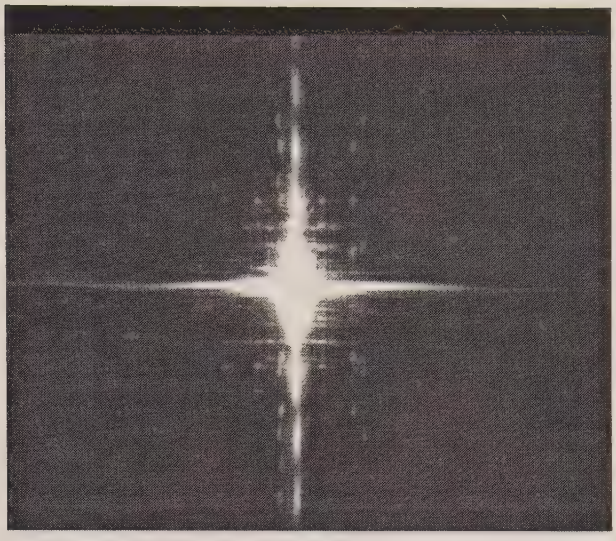


Fig. 3 Correlation plane corresponding to target in Fig. 2

The next step is to use polarization enhancement during image capturing. Fig. 4(a) shows a polarization-enhanced image of the same case. Fig. 4(a) is generated using Eq. (3). It corresponds to S_1 image in the SVI. Fig. 4(b) is a plot of correlation output and Fig. 4(c) is a two-dimensional plot of Fig. 4(b).

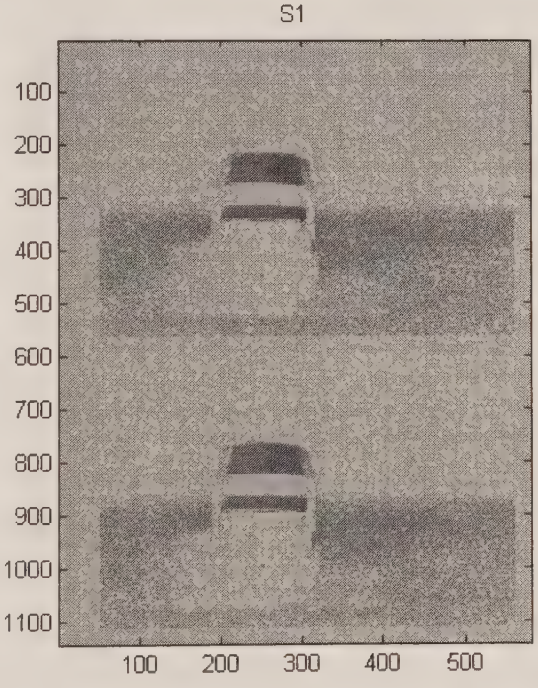


Fig. 4(a): S_1 image of Fig. 2(a) generated using Eq. (3).

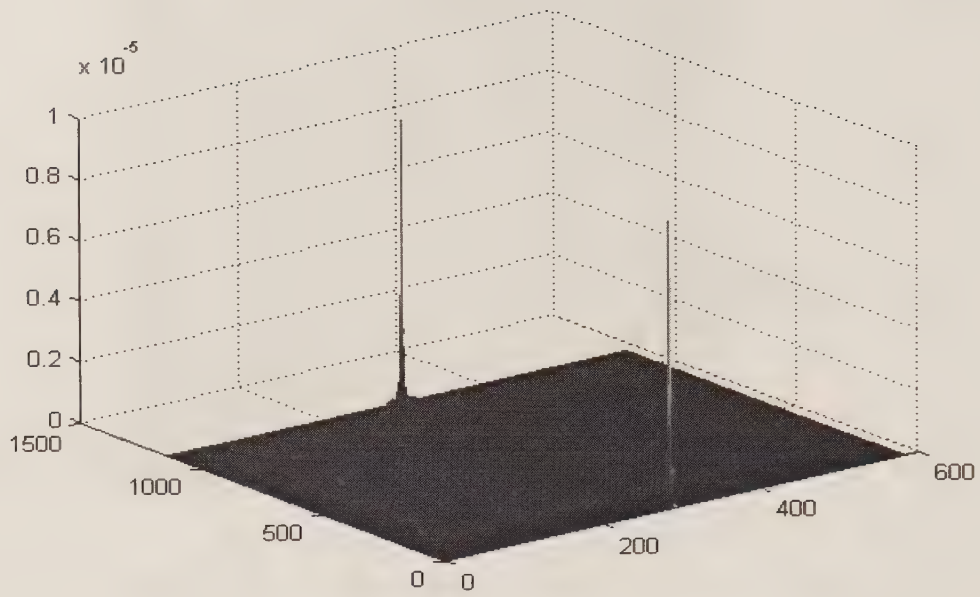


Fig. 4(b): Correlation output

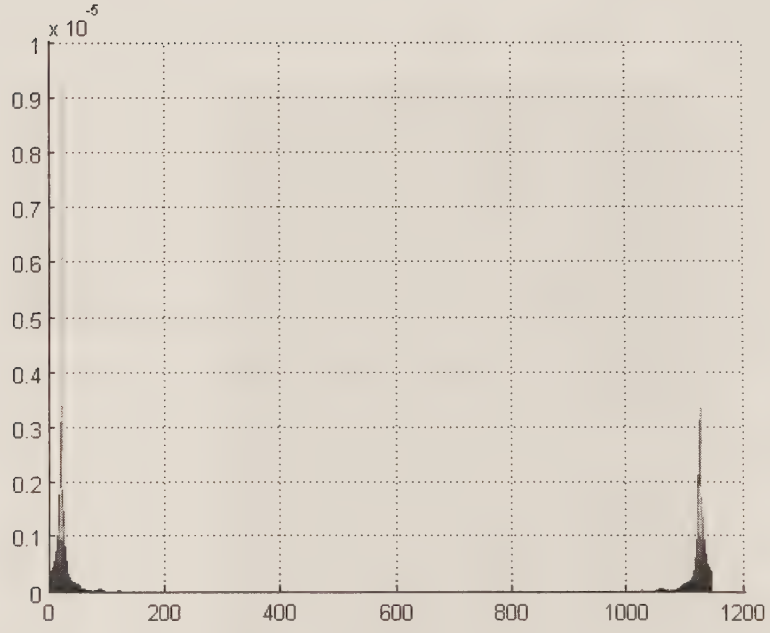


Fig. 4(c): 2-dimenstional view of Fig. 4(b).

The calculated performance parameters for this case are: P ≈ $94x10^{-7}$, PSR ≈ 43.64, and PCE ≈ 2.9%. Compared to Fig. 2(c), Fig. 4(c) shows approximately fifteen times increase in P, more than four times in PSR and more than seven times increase in PCE. Compared to Fig. 2(c) Fig. 4(c) contains narrower, sharper and higher correlation peaks. This significant improvement in correlation parameters obtained from computer simulation results are further supported by the following experimental results.

Fig. 5 shows the experimental correlation output corresponding to this case. Fig. 5 unmistakably shows the location of the target as bright spots in the first and third quadrant of the correlation output. In addition compared to Fig. 4 we notice a significant reduction in source noise around the dc terms as shown in Fig. 5. It is evident in this case that the probability of target misses and false alarms has been remarkably reduced.

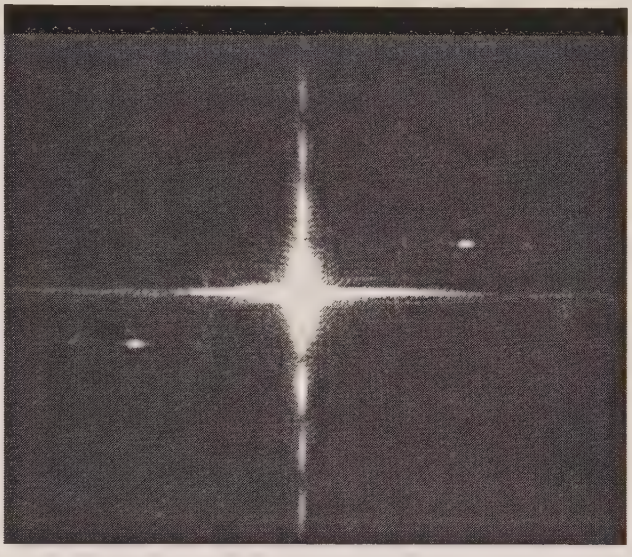


Fig. 5 Correlation plane showing the strong peaks corresponding to the target, subjected to polarization enhancement.

We further note that the polarization-enhanced images captured were further subjected to digital enhancement techniques before applied to the FJTC. The results showed further improvements in correlation parameters. Analyses of the new results were not included here, in order to remain within the space limitation of this manuscript.

5. Conclusions

We introduced novel optoelectronic passive visible polarimetric imaging target detection system using polarization-enhanced images as inputs to the FJTC. Our simulated and experimental results showed significant improvements in the correlation output parameters. The proposed technique showed improved target detection and discrimination capabilities that lead to remarkable decrease in target misses and false alarm rates.

REFERENCES

1. Joseph L. Pezzaniti and Russell A. Chipman, "Imaging Polarimeters for Optical Metrology", SPIE Vol. 1317, p. 280-294 (1990).
2. Firooz Sadjadi and Cornell Chun, "Machine Recognition of Objects using IR Polarimetery", SPIE Vol. 2756 (1996).
3. P. De'mencean , S. Breugnot and L. Collot," Polarization Diversity Active Imaging", SPIE Vol. 3880, p. 284-291 (1998).
4. Gregory P. Nordin, Jeffrey T. Meier, Panfilo C. Deguzman, and Michael W. Jones, " Micropolarizer array for infrared imaging polarimetry", *J. Opt. Soc. Am.* A, Vol. 16, No. 5 (1999).
5. Se'bastien Breugnot and Phillippe Cle'menceau, "Modeling and performance of a polarization active imager at λ= 806 nm", *Opt. Eng.* 39 (10), p. 2681-2688 (2000).
6. M. J. Duggin and R. S. Loe, "Calibration and exploitation of a narrow-band imaging polarimeter", *Opt. Eng.* 41(5), p. 1039-1047 (2002).
7. R. M. Azzam, "Potopolarimetric measurement of the Mueller matrix by Fourier analysis of single detected signal", *Optics Letters*, **48**, p. 148-150 (1990).
8. D. H. Godstein, R. A. Chipman, and D. B. Chenault, "Infrared Spectropolarimeters", *Opt. Eng.* 28, p. 120-125 (1989).
9. J. Wang and B. Javidi, "Multiobject detection using the binary joint transform correlator with different types of thresholding methods," *Opt. Eng.* 33, p. 1793-1804 (1994).
10. M. Deutsch, J. Garcia, and D. Mendlovic, "Multichannel single-output color pattern recognition by use of a joint-transform correlator", *Appl. Opt.* 35, p. 6976-6982 (1996).
11. M. S. Alam and M. A. Karim, "Fringe-adjusted joint transform correlator", *Appl. Opt.* **32**, p. 4344-4350 (1993).
12. M. S. Alam, "Performance of Joint Transform Correlation Under Varying Illumination and Noise," *Journal of Optics and Laser Technology*, Vol. 27, p. 361-368 (1995).
13. B. V. K. Vijaya Kumar and L. Hassebrook, "Performance measures for correlation filters", Applied-Optics. vol. 29, no.20; p. 2997-3006 (1990).
14. J. L. Horner, "Metrics for assessing pattern recognition performance", Appl. Opt. 31, p. 165–166 (1992).
15. M.A. Philips, S.R.F. Sims, "A signal to clutter measure for ATR performance comparison ", *In proceedings of SPIE Automatic Target Recognition VII,* p. 74-81, Vol. 3069 (1997).

Passive Optical Methods for Helmet Tracking in Aircraft

Patrick J. Vitarius and Don A. Gregory
University of Alabama in Huntsville
Huntsville, AL 35899

Denise Morell
FastMetrix
4040 Memorial Parkway SW
Huntsville, AL 35802

Introduction

Helmet tracking refers to the practice of remotely sensing the position and orientation of an aircraft pilot's helmet with reference to the cockpit. Compared to the more established mechanical and magnetic methods, optical methods are in their infancy. Any system that will aid the military pilot in directing fire or sensors or aid in the stabilization of imagery that is captured by the aircraft's sensors will substantially increase the pilot's effectiveness. An additional, and important, benefit of optical methods is that they avoid the undesirable secondary effects intrinsic in magnetic and mechanical systems.

Fundamental physics guarantees that, given the opportunity, time-varying magnetic fields will induce currents in any conductor that is nearby[1]. It is extremely difficult to shield against stray magnetic fields. The small currents induced in electronic equipment can be a serious problem and lead to other vehicle system failures. Mechanical methods typically involve coupling the pilot's helmet to a stationary point in the cockpit through the use of rigid rods, joints, and pistons. Such configurations are undesirable from the standpoint of safety, as they restrict the movement of the pilot as well as present potentially dangerous cockpit debris in the event of an accident.

Various researchers have examined optical methods for determining position and orientation. Virtual reality applications where the computer must 'know' the position and orientation of the user's head and body are one field where the method has been exploited to a large degree. Visidyne's HelmeTracker[2] uses time of flight algorithms, and Spica's Helmet Tracker[3] uses direct view. Cumulus Corporation's Kentron[4] is also developing an optical tracking system for helmets. Much of the current research in the field is based on paradigms developed for magnetic systems or acoustic (ultrasound) techniques.

Several optical techniques are described for determining the orientation and position of a pilot's helmet. These techniques are all passive in that they require no input from the pilot except perhaps for an initial calibration, which would likely be valid for a particular pilot for any subsequent flight. Eight distinct methods are presented and the advantages and disadvantages of each are described. All of the methods are based on sound optical principles, but with current technology considerations in mind, it is likely that some will be better than others at meeting the required criteria.[5]

Optical Pattern Recognition XV, edited by David P. Casasent, Tien-Hsin Chao, Proc. of SPIE Vol. 5437 (SPIE, Bellingham, WA, 2004) · 0277-786X/04/$15 · doi: 10.1117/12.543243

The problem of determining the position and orientation of a pilot's helmet can be thought of as determining the position and orientation of a vector that starts at the center of the pilot's head and points forward. If we take this vector to be of a fixed (arbitary) length, the problem reduces to the determination of five independent variables.

Our general approach is to use any of a variety of active or passive emitters on the helmet (hereafter 'sources') and any of a variety of optical measurement devices located off of the helmet (hereafter 'detectors').

Figure 1. Determining the position and orientation of the pilot's helmet amounts to determining the vector that starts at the center of the pilot's head and points forward.

Each detector-source pair will produce one or more data points. For example, an LED-photodiode pair will produce a single data point, the voltage across the photodiode. A video camera pointed at a small dot on the side of the helmet will ultimately produce two data points, the xy-coordinates of the image of the spot in the image plane. Issues of redundancy aside, the number of data points obtained must be greater than five in order to determine the five independent variables that determine helmet position and orientation.

A problem that increases the number of detector-source pairs is that of cosine ambiguity. It is common to many of the techniques described below that the value measured depends not on the orientation angle but on the cosine of the orientation angle. Cosine is an even function, this means that each value obtained for the orientation angle will have a sign ambiguity. In order to resolve this ambiguity, additional detector-source pairs must be added. Note that in many applications, it is possible to add just a detector or just a source in order to increase the number of pairs.

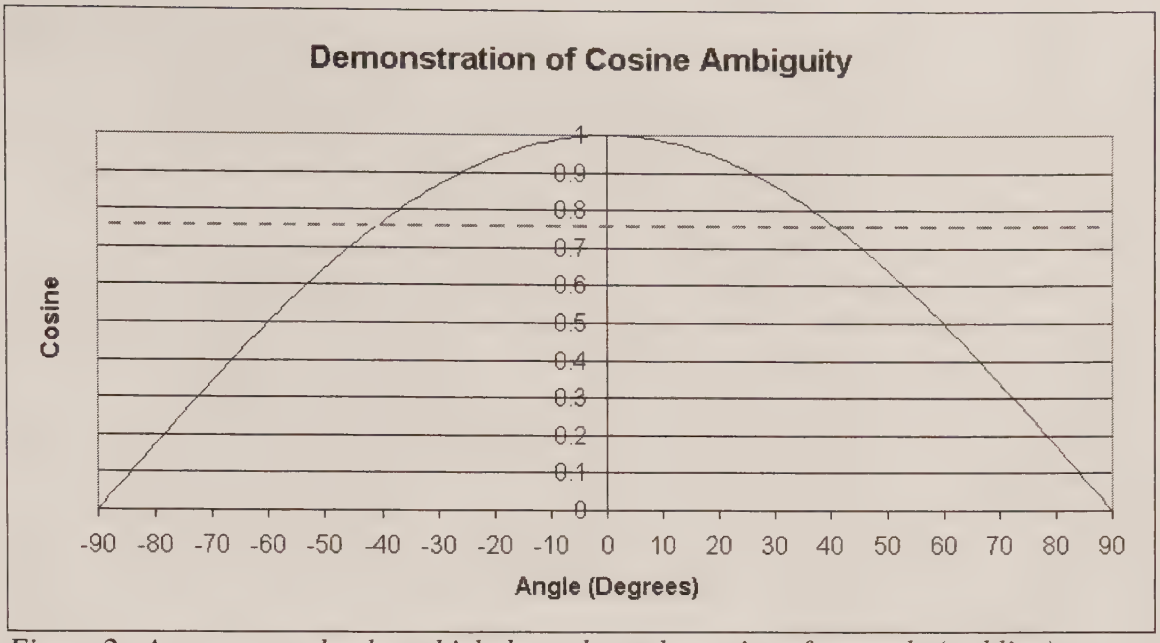


Figure 2. Any measured value which depends on the cosine of an angle (red line) corresponds to two possible values of the actual angle.

Another artifact of cosine dependency is non-linearity of the response. The sensitivity of the detector—the change in the measurable quantity for a given change in helmet orientation angle—is strongly angle dependent, and falls to zero for small angles. Additionally, the response of the detector such as a photodiode is angle dependent, falling to zero for angles near 90°. A similar fall-off may be noted in optical techniques that rely on image processing: at larger angles, two points on the helmet separated by a certain finite distance will form of an image of two points at progressively smaller separations, so that the precision of the value obtained for angle becomes limited by the resolution of the imaging device.

Thus, depending on the technique, an array of sensors may be desirable so that one sensor is always in quadrature, experiencing an optimal balance of sensitivity and response.

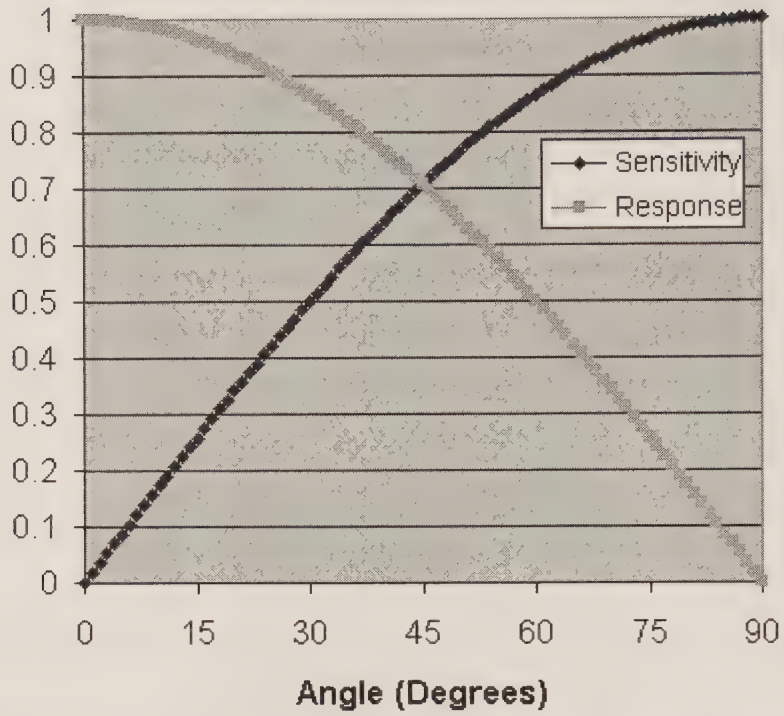


Figure 3. While sensitivity increases with increasing angle, response decreases. Thus the optimal angles for a balance of sensitivity and response is around 45°.

Direct View

The first technique discussed is direct view, which is directly viewing the pilot's helmet with a high-resolution camera. This involves treating the helmet, or objects painted on the helmet, as passive reflectors of ambient light. Lighting conditions are obviously a factor here; a system that works well in daylight will not work as well in nighttime conditions. Flood illuminating the cockpit with white light at night may not be an option, as it would limit the pilot's ability to see through the cockpit. Flood illumination with infrared may be an option. In fact common CCD cameras will work into the infrared range to some degree.[6] Image jitter caused by the vibration of the camera and the vibration of the pilot will also be a problem if some sort of image stabilization or registration is not used. It is also possible to flood illuminate with several narrow-band sources, e.g., diodes, as a method allowing results comparable to those obtained by flood illumination at a fraction of the wattage. A highly retro-reflective paint similar to that used for marking street dividing lines can be used to direct most of the flood illumination back toward the camera to avoid scattered light problems. This has been successfully used recently in determining the optical figure of large inflatable optical elements. A series of small dots were painted on the unused surface and white light from the direction of the cameras was used as the illumination source. The reflection from the dots created a grid that could be analyzed to determine the optical shape of the element.[7]

However, there are direct view techniques that do not require imaging the entire helmet. A simple bar chart—a Ronchi ruling—painted in a known location on the helmet provides information about rotation in a direction orthogonal to the bars. As the bar

pattern rotates the bars appear closer together to a camera having a flat imager, as most do. It is a fast digital operation to calculate the separation and to determine the tilt angle. The apparent separation between bars l' is related to the orientation angle θ by the relation

$$l = l' \cos\theta \,. \tag{1}$$

Hence each Ronchi ruling provides information along one axis of orientation, with a cosine ambiguity. An array of three Ronchi rulings, all possibly observed with the same camera, would be sufficient to uniquely determine the orientation of the helmet. Computationally, this is a very simple calculation so that the update rate can approach the limit of the number of frames per second captured by the camera.

Helmet position can also be determined by direct view of point sources on the helmet. By calculating positions of images on the image plane, actual location can be calculated from purely geometric concepts. The same camera that makes the orientation measurements can obtain information regarding the x'y' coordinates of the point sources, where x' and y' are the axes in the image plane of the camera. However, to calculate x and y from x' and y', it is necessary to know z, the distance of the helmet from the camera. A second camera mounted along another viewing axis can provide this information, as can a second type of image whose Fourier transform is highly scale (range) dependent. An example of an image of this type is a series of concentric rings. Alternatively, it may be possible to do some image algebra on the entire scene (a derivative operator for example) to differentiate the edges of the helmet as seen by the camera. By knowing the apparent radius of the (nearly spherical) helmet, it would be possible to obtain the distance to the helmet from the camera. Again, this is not a complicated operation and one that is routinely done in the movie industry to model the movements of a "stick figure" created by a series of LED's or reflectors placed on a stunt actor's clothing.[8]

Color Mixing

A second approach involves the placement of 3 (or more) narrow band light emitting diodes (LED's), or the same number of broadband LED's with narrow band filters, on the helmet. These devices are currently available, and together with a battery that will power the LED for six or more hours, will fit into a package no larger than a stick of gum.[9] The package can be made even smaller if the LED's can be powered by the helmet's power supply. Only a few milliamps are required at 5 – 6 volts DC.

Three (or more) detectors are matched to the diodes' wavelengths, and placed in the cockpit at predetermined locations such that the percentage of light of wavelength λ received by each detector determines the position of the helmet. The system is calibrated by having the pilot look at a series of reference points inside the cockpit, providing an intensity reference for each of the detectors. This only takes a few seconds and can be repeated at any time. If each diode source is configured to behave as a Lambertian emitter, then the angular dependence of the power emitted is

$$P = P_o \cos\phi, \tag{2}$$

where ϕ is a measure of angular deviance from normal incidence on the detector. Thus each detector can calculate the angle of separation between the forward-pointing vector and another, fixed vector. Once the angle of separation is known for at least three distinct vectors, the orientation of the helmet can be uniquely determined. In the quadrature range, a change in angle of 10mR corresponds to a change in power level of about one percent peak power, well within the sensitivity of modern photodiodes.[10]

A related method involves the placement of different color scenes or dots on the helmet and them viewing the scene in white light. The principle is the same- using the percentage of color mixing to indicate the position of the target. This implementation can be demonstrated in the Phase 1 effort using common CCD cameras and white light illumination. This technique will probably not be acceptable in application since it would require illuminating the whole cockpit or at least the helmet with white light. However, it might be possible to create an analog in the infrared or even ultraviolet region, where the targets on the helmet are made of material that reflects well in these regions and the camera employs filters to distinguish the "colors" reflected.

Optical Correlation

Optical correlation is a powerful technique for comparing two or more scenes. It is possible to construct a holographic system based on matched filters, coherent illumination, and spatial light modulators that will recognize any scene and discriminate between similar scenes.[11] This capability has also been utilized, with specially constructed scenes, for determining orientation. This technique was proposed in the late 80's by researchers at the US Army Missile Command at Redstone Arsenal, Alabama and the Johnson Space Center for docking spacecraft oriented arbitrarily with respect to each other.

Digital versions of the optical correlation technique have recently become powerful enough to handle simple scenes. NASA's 2003 Commercial Invention of the Year Award was presented to a Marshall Space Flight Center scientist for employing the technique in image stabilization.[12] The same ideas can be used, with a sufficiently complex scene (as regarding spatial frequency) to determine orientation. This would require perhaps three scenes of different spatial frequencies, placed in different locations on the helmet. A simple camera placed in the cockpit would collect a continuous image of the helmet. The scene would be captured and continuously Fourier transformed. The spectra would contain a mixture of spatial frequencies determined by how much of each scene is observed, which depends on the angular orientation of the camera.

This idea can be taken one step further and employ a matched filter of each of the scenes. The strength of the correlation signal and its spatial location (translational invariance) then determines the orientation of the helmet, as well as the location of the helmet in the unprimed coordinate system.[13]

Diffraction Gratings

This method involves placing small thin diffraction gratings (amplitude or phase) on the helmet itself, then flood illuminating the helmet with light from a broad band LED or three (or more) narrow band LED's. This requires detectors (probably one for each grating) strategically placed in locations in the cockpit. Diffraction gratings are extremely wavelength and angle sensitive, which can be exploited for accuracy in positioning measurements. The grating equation is:

$$n\lambda = d\sin\theta\,.^{14} \tag{3}$$

As the helmet (and gratings) are tilted and rotated, the detector perceives a shift in grating spacing according to:

$$d = d_o \cos\theta\,, \tag{4}$$

where d_o is the true spacing as seen from the normal, which would possibly be the calibration point. Note that there is a cosine ambiguity in this technique as well. This problem could be resolved using multiple grating/sensor pairs. Locating the centroid itself, however, is another problem. This will likely rely on information gained from the image of the helmet itself. It is not a complicated matter to store, digitally, the possible locations of the centroid and then determine the actual, live location by using a look-up table and key characteristics of the helmet itself, such as common markings (or shapes).

Time of Flight

A technique based on the time of flight of an optical pulse to a known target and back to a detector tuned to a specific wavelength is a realistic choice as well. This is essentially laser radar and has proven to be a valuable technology over the years.[15] A short pulse of light is emitted—the shorter the better for position sensing—and travels to the target. A detector placed very near the source determines the round trip time or "time of flight". This was actually a commercially available technology in the 70's and employed fiber optic bundles to transmit and receive the optical pulses.[16] It was possible at that time to distinguish displacements of a few microns. The same idea can be applied to determining the position of a pilot's helmet.[17] Three (or more) "targets" can be placed on the helmet and reflect the incident pulses. In this case the targets will be small reflectors (or corner-cube retroreflectors) that will direct the incident light back toward the source and detectors. The light sources themselves can be practically anything that can deliver short pulses of illumination. LED's can be pulsed at multi-kilohertz rates quite easily and can be found in a wide range of wavelengths. Three sources (with integral detectors) can be used to determine θ and ϕ and an additional source can provide redundant information as well as displacement from a centroid location (determined by an initial calibration). The incident and received pulses can be routed using optical fibers, allowing the sources and detectors to be located almost anywhere. With suitable optics and retroreflectors, it is possible to obtain a measurement at any rotational and translational position.

Intensity Gradient

Another passive optical technique involves creating a visual gradient across the surface of the helmet—for example, the helmet's color could fade gradually from red on the left side to white on the right side. By fixing a camera or filtered photodiode at a narrow viewing angle, the direction the helmet is facing can be inferred from the intensity of reflected light in that wavelength band. To accommodate multiple axes of rotation, additional gradients could be superimposed on the first, perhaps using a modified Bayer filter.[18] An advantage of this method is that it requires very little computational power. Each sensor is trained at a fixed location and measures the intensity within a certain frequency band; this intensity depends in a predetermined manner on the orientation angle of the helmet. There is no cosine ambiguity because the gradient pattern is asymmetric. A disadvantage of this method is that it is very dependent on ambient lighting conditions. It may be possible to use reference sensors trained on stationary objects to compensate for changes in ambient lighting, or perhaps utilize a wide bandwidth, low intensity illumination so that ambient lighting would not be critical.

Fiber Optic Gyroscopes

Fiber optic gyroscopes are currently used for many applications from missile guidance to measuring stresses in buildings, and robotics. In a ten-year-old article[19] these devices were used to determine head movement in vestibular reflex measurements. The same general idea can be applied to determine the movement of the helmet the head could be wearing. These researchers found that the fiber optic gyros provided much better performance than the other methods used at the time. The methods mentioned included magnetic pick-up technology similar to that used in the Apache helicopter. Eye movements were also recorded in the study using electrooculography. The findings concluded that the real improvement was in the signal to noise ratio. The fiber optic system was simple to install on the harness worn on the head and easily exceeded the requirements of power, cost, and simplicity of signal routing. Signals were sent to external instrumentation using infrared diodes and detectors at a distance of more than a meter away from the LED's. In the last ten years a great deal of improvement has been made in fiber optic gyro technology and in LED's. This technique does not determine absolute position of the head in a three dimensional space but with the addition of perhaps one more LED, this might also be accomplished.

Spectral Polarization

Perhaps the best passive technique will combine two concepts (or more). Polarization encoding is a powerful tool that can be utilized quite easily. It is a simple matter to include a thin sheet polarizer over any of the LED's used. The same type of polarizer can be utilized over the detector(s) used to collect the optical information. The sensitivity of polarization to angle is well known and expressed as Malus's law:[20]

$$I = I_o \cos^2 \theta, \tag{5}$$

where θ is the angle between the polarization state of the encoded source and the transmission axis of the polarizer analyzer affixed to the detector. I is the detected irradiance and I_o is the incident irradiance. If each of the diode sources is not only spectrally unique but also polarization unique, then both characteristics can be used to determine the angular orientation of the helmet. This adds practically nothing to the cost, weight, or complexity of the system, yet adds considerable sensitivity, as seen in the above equation where the irradiance is now a function of the square of the cosine of the angle, not simply cosine of the angle as in previous concepts. If this new sensitivity is then combined with one of the other techniques such as time of flight then it seems that the angular orientation as well as the absolute distance from the source are known to a high degree of accuracy. This requires a pulsed LED, which is not an added complexity at all. In fact most diodes work well in a pulsed mode. Polarization encoding has not been explored at all by any of the other researchers in the area and will likely prove to be the most sensitive method for determining the orientation of the pilot's helmet. If nothing else, the use of polarization can add a redundancy to the system that is not present at the current time. A different polarizer orientation can simply be added to each spectrally independent LED. Then each detector, which is already spectrally sensitive (discussed in a previous method), now becomes polarization sensitive as well. Polarization encoding can be combined with most of the techniques discussed above to improve sensitivity.

Conclusion

The methods outlined above demonstrate that for this particular tracking problem, there are numerous solutions, and more than one may be 'correct.' The task that remains for the system designer is to decide which is the most correct for the current application. The solution that provides the desired degree of accuracy, that can also be built and operated within acceptable cost limits, and that can be sufficiently ruggedized, miniaturized, and otherwise made to conform with the specifications of the ultimate application has yet to be determined.

---

[1] Rash, C., "Visual Coupling", http://www.usaarl.army.mil/hmdbook/cp_004.htm and Rash, C., Mozo, B., McLean, W., McEntire, B., Haley, J., Licina, J., and Richardson, L., "Assessment Methodology for Integrated Helmet and Display Systems in Rotary-Wing Aircraft", Fort Rucker, AL; U.S. Army Aeromedical Research Laboratory, USAARL Report Number 96-1

[2] The Visidyne HelmeTracker, http://www.visidyne.com/helmetracker.htm

[3] The Spica Helmet Tracker. http://www.spicatek.com/Support/HT.pdf

[4] Kentron's ARCHER Tracker.http://www.cumulus.co.za/Default.asp?Type=Tracker

[5] Comanche Production Performance Weapon System Specification, paragraph 3.3.7.1.2.1

[6] Indigo Systems, Omega infrared CCD camera, *Photonics Spectra*, June, 2003

4 Pearson, J., and Unroe, M., Precision Tooling for Membrane Concentrators, *AIAA Joint Propulsion Conference,* Huntsville, AL, July, 2003, paper 5174
8 Keyton, H., "Industrial Light and Magic's Seth Rosenthal on Motion Capture", *Animation World Magazine*, Issue 3.11, February, 1999
9 *Laser Focus World Buyers Guide*, 2003, Pennwell Publications
10 Silicon Detectors, Edmund Industrial Optics Catalog, 2003, p. 198
11 Gregory, D. and Kirsch, J., *Applied Optics*, Vol. 31, p. 163, 1992
12 Rosenberg, B., "NASA Technology Sharpens Images With Jitter", *Aviation Week and Space Technology*, May 19, 2003, p.61
13 Goodman, J., *Fourier Optics*, McGraw-Hill publisher, 1968, p. 9
14 Hecht, E., Optics, Addison-Wessley publisher, 1987, p. 333
15 Becherer, R. (Ed.), Laser Radar V., Proc. SPIE, Vol. 1222, 1990
16 Mechanical Technology Inc., Latham, NY
17 Blais, F., Beraldin, A., and El-Hakim, S., "Range Error Analysis of an Integrated Time of Flight Triangulation, and Photogrammetric 3D Laser Scanning System", *Laser Radar Technology and Applications V.*, Proc. SPIE, Aerosense 2000, Vol. 4035
18 Jacobson, R., Ray, S., Attridge, G., and Axford, N., *The Manual of Photography*, 9th Edition, Focal Press, Division of Elsevier Publications, 2000
19 Donati, S., Annovazzi-Lodi, V., Bottazzi, L., and Zambarbieri, D., "Pickup of Head Movement in Vestibular Reflex Experiments with an Optical Fiber Gyroscope", *IEEE J. Sel. Top. Quant. Elect.*, Vol. 2, No. 4, Dec., 1996, p. 890
20 Pedrotti, F., and Pedrotti, L., *Introduction to Optics*, Prentice Hall, 1993 (299)

Neural network tracking and extension of positive tracking periods

Jay C. Hanan,[1*] Tien-Hsin Chao,[1] Pierre Moreels[2]

[1]Bio-Inspired Technologies and Systems, Jet Propulsion Laboratory, CA, USA
[2]Electrical Engineering, California Institute of Technology, CA, USA

ABSTRACT

Feature detectors have been considered for the role of supplying additional information to a neural network tracker. The feature detector focuses on areas of the image with significant information. Basically, if a picture says a thousand words, the feature detectors are looking for the key phrases (keypoints). These keypoints are rotationally invariant and may be matched across frames. Application of these advanced feature detectors to the neural network tracking system at JPL has promising potential. As part of an ongoing program, an advanced feature detector was tested for augmentation of a neural network based tracker. The advance feature detector extended tracking periods in test sequences including aircraft tracking, rover tracking, and simulated Martian landing. Future directions of research are also discussed.

Keywords: autonomous tracking, feature extraction, feature detector, neural network, target recognition, data reduction.

1. INTRODUCTION

Automatic target tracking technology has a wide variety of applications. Tracking technology has been researched for use with navigation of Martian landers, rovers, or on Earth for avionics testing or various military applications. Domestic applications abound and a successful technology which is capable of robust application to a variety of scenarios with little expert technical attention is still under investigation. Specific applications often have solutions where a peculiar feature may be available, but that feature is not expected applicable to the general tracking problem. In order to obtain a general purpose tracking tool, a neural network method was employed. The goal of such a tool was to simplify the necessary inputs given by the operator; thus, a general robust, yet simple method was sought.

A previous examination of a neural network based tracking system showed the potential for improved tracking periods (the number of frames where tracking remains precisely on target).[1] However, information input to the neural network was limited, so that potentially useful image information was ignored—primarily in the interest of computational efficiency. Here we reexamine this technology with a consideration for conserving potentially useful information available to the system. Advanced feature detectors were applied to the entire frame. While tracking of only a small component in the frame may be sought, this additional information provides data such as camera pose, frame movement, object rotation, and scale. The advanced feature detector focuses on areas of the image with significant information. Basically, if a picture says a thousand words, the feature detectors are looking for the key phrases (keypoints). These keypoints are rotationally invariant and may be matched across frames. Application of these advanced feature detectors to the neural network tracking system at JPL has promising potential. Some tests of this technology with example applications are described below.

2. METHOLOLOGY

An analysis of the tracking results reported previously suggests the stable tracking period for the examined sequence can be extended beyond 100 frames, if sufficient information can be input to the system to correct the instabilities.[1] A proof may be sought from the results of human tracking operators, traditionally used to mark object trajectories. The human operator uses additional information beyond the tagged area of interest to maintain awareness of the target. To input this

* Corresponding author; electronic mail: Jay.C.Hanan@jpl.nasa.gov

Optical Pattern Recognition XV, edited by David P. Casasent, Tien-Hsin Chao, Proc. of SPIE Vol. 5437 (SPIE, Bellingham, WA, 2004) · 0277-786X/04/$15 · doi: 10.1117/12.548081

additional information into the automatic tracking software, rotationally invariant features across the image are useful. Furthermore, these features must be matched between sequential frames so that neurons may be trained based on similar objects. However, such an operation ordinarily would cost excessive computing resources. Thus, a compromise considering resources and accuracy must be determined.

Feature detectors have been considered for the role of supplying additional information to the neural network tracker. Feature detectors have been employed for some time. Lowe presents a recent feature detector that has been used with success.[2] The key ideas behind features detectors are simple. Of course, the best description of the image is the image itself. However, it takes considerable of time and space, so we want to store only small collections of data about an image. Homogenous zones contain little information, and more important, any zone from a homogenous region looks exactly the same as any other, giving poor localization (if you cut a small patch from a homogenous region, you have no way of telling where this patch has been cut). Many detectors look at places where there is a high gradient. Gradients are typically computed with filters such as:

$$\begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}$$

This "Sobel" filter does some smoothing to remove noise, and at the same time calculates the gradient at the center point. The gradient in one direction is not enough: if you identify the edge of a table, you have good localization in the direction perpendicular to the edge, but poor localization in the direction of the edge. Thus, we must make use of derivatives in both x and y. One detector that was developed 10 years ago and still used frequently is the Harris detector. The Harris detector takes the matrix

$$M = \begin{bmatrix} \left(\frac{d}{dx}I\right)^2 & \left(\frac{d}{dx}I\right)\cdot\left(\frac{d}{dy}I\right) \\ \left(\frac{d}{dx}I\right)\cdot\left(\frac{d}{dy}I\right) & \left(\frac{d}{dy}I\right)^2 \end{bmatrix}$$

and then takes the local extrema of

$$R = \det(M) - k(\mathrm{tr}(M))^2$$

There are other feature detector methods: the Foerstner detector and the Hessian detector, which are based on the second derivatives. The detector used by Lowe is based on filtering with a difference of gaussians,[2] similar to a laplacian filter—again second derivatives. Recently, the detections have been applied in multiscale: first one detection is performed, then resample at a lower resolution and perform detection again. This is in order to match features even if there was a change of scale between views.

It is difficult to objectively determine which feature detector is 'the best'. It is in large part application-dependent.[3] Beyond an extension of previous work, in investigating these advanced feature detectors, we applied the Hessian detector to a series of simulated landing images for a Mars entry, decent, and landing sequence. In addition, we tested the neural network tracking system as part of a path planning test bed for extended tracking of rovers and obstacles over terrain. Results from these tests are given below.

3. PRELIMINARY RESULTS

3.1 Aircraft Tracking

Consistent with previously reported results,[1] the additional information provided by advanced feature detectors can extend the stable tracking period the length of the sequences (>300 frames compared to the 100+ reported earlier). However, optimal automatic selection of scale and feature density is still under investigation. The system was tested on the sequences analyzed previously along with a rocket sequence depicted in Figure 1. The previously tested sequences provide a test of rotation and small changes in scale. The rocket sequence provided a test of more extreme changes in scale along with some cloud obscurations. The addition of occlusions provided a good opportunity to test the advanced feature detector. Advantages in target reacquisition and frame movement tracking were observed. As with the previously tested sequences, tracking periods improved with the input of additional information. The rocket sequence was positively tracked with periodic loss of target associated with cloud cover.

Figure 1 An early frame in a sequence where successful tracking of a rocket through large changes in scale and with the presence of obscurations such as clouds was demonstrated.

3.2 Rover Tracking

The advanced feature detectors were unnecessary for positive indefinite tracking of the rover in the test bed set up (Figure 2 A). Tracking of the rover across the test bed was successful on all trials.[4] Variation in lighting and rover travel rate were tested against the system. Rover tracking was consistent and continuous up to the maximum rate of travel available by the system (~0.15 m/s).

Along with motion tracking of the rover science craft vehicle, science targets (represented by blocks in Figure 2 A) must also be identified. While this is a simple segmentation task for the blocks, a more advanced recognition system will be employed for field tests. Feature detectors were tested and found useful for applications requiring identification of natural science targets (Figure 2 B), such as the rocks expected in field application. The features on the rock in Figure 2 B were tracked through a 360 degree rotation. Incorporation of the feature detector to the rover system is part of the ongoing work of this project.

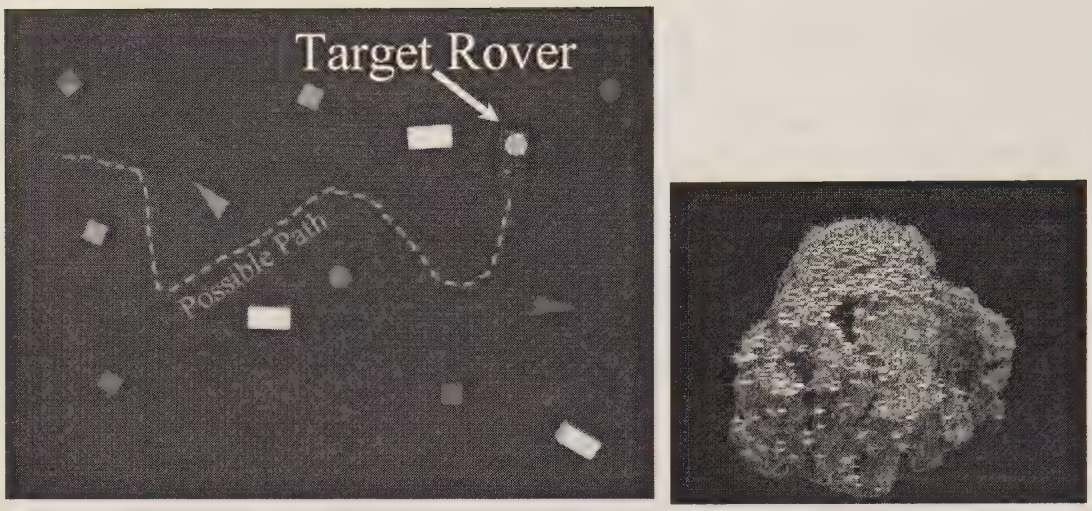


Figure 2 A) Test bed for rover path planning and obstacle avoidance tests. B) Feature detector used to identify "rock" target.

3.3 Mars Landing

The landing sequence provided a good test case for use of the advanced feature detector. The change of scale from high-altitude to surface provided a difficult case for the neural network tracker. Tracking becomes less stable since the landing site is of low contrast and the scale and rotation changes are significant throughout the sequence. Here, since the object of interest is a safe landing site of low contrast, tracking of "off-target" objects throughout the image was necessary to maintain consistent tracking. Consistent recognition of craters between images was achieved using the advanced feature detector (see Figure 3). The simulation indicates the tool could perform tracking throughout the landing sequence. Testing of the optimal interconnection between the advanced feature detector and the neural net target tracking system remains in progress.

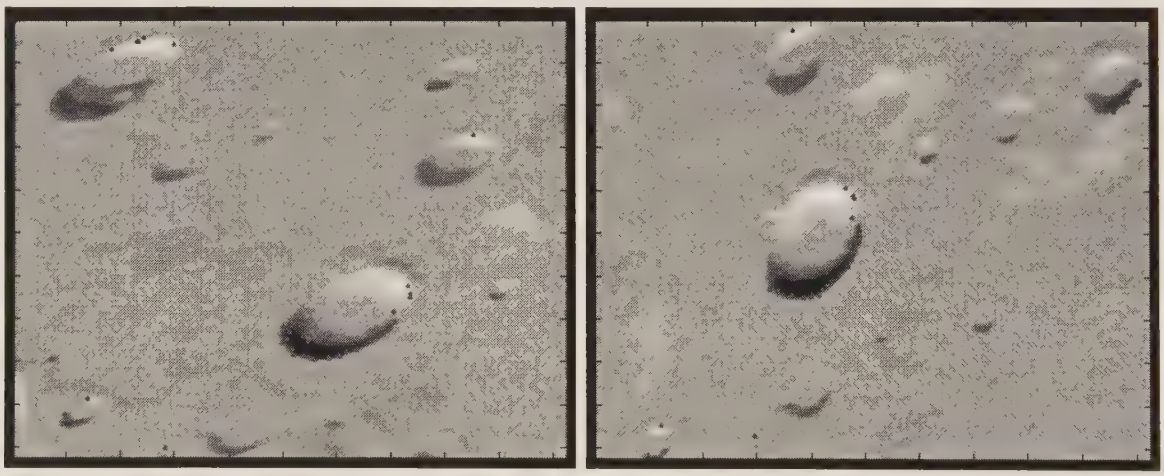


Figure 3 Extraction of two frame showing rotation and translation of the simulated Martian surface.[5] The craters provide sufficient contrast to be recognized across frames using the advance feature detector alone. Small blue spots mark areas where the advanced feature detector found a match.

4. SUMARY

An advanced feature detector has been tested for augmentation of the neural network tracker. It is clear a second level of information extends tracking periods as evidenced by the tracking of the entire aircraft flight sequence as compared to the portion of the sequence reported previously.[1] However, the best method to incorporate this second level of information is still under investigation. Additional results will be presented in Optical Pattern Recognition XV.

Potential improvements may be realized by extending the number of frames over which features are compared, particularly where clouds may temporarily obscure the target. An automated connection between the confidence level given by the neural network and the level to which dependence on the advanced feature detector is a necessary first step. Planned field application to a longer sequence of up to 1 million frames will provide opportunities for testing further extensions of the tracking periods.

A case of indefinite positive tracking was confirmed on a simple target consisting of a rover for a path planning application at JPL. In this case the scale remained constant, but rotation was often extreme. Future application of the advanced feature detector to recognition of science targets and tracking of "natural" rock targets over multiple frames is also in progress. A final application of the advanced feature detector to simulated Mars landing images provided an efficient use of the method. The addition of the advanced feature detector significantly increased the tracking period for this sequence over that which was possible with solely the neural network tracking method. Additional tests of the method to real mars landing data will also be presented.

ACKNOWLEDGEMENTS

The research described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

REFERENCES

[1] J. C. Hanan, H. Zhou, T. Chao, "Precision of a radial basis function neural network tracking method," *SPIE, Optical Pattern Recognition XIV*, 5106 (2003).

[2] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," accepted for publication: *International Journal of Computer Vision*, (2004).

[3] C. Schmid, R. Mohr, and C. Bauckhage, "Comparing and Evaluating Interest Points," *IEEE Proc. Int'l Conf. Computer Vision*, pp. 230-235, (1998).

[4] W. Fink, J. C. Hanan, T.-H. Chao, M. A. Tarbell, "Field-Deployable Integrated Air-Ground Multi-Agent Autonomous Remote Planetary Surface Exploration." *NASA NTR-40428* (2003).

[5] N. J. Villaume, B. J. Martin "Mars entry, decent, and landing image sequence simulation" *unpublished data* (2003).

Hand Motion Recovery Based on Skin Color Segments and Motion Parameter Estimation *

S. Chang, L. Elkins, Z. Wang, T. Zhou, S. Foo, A. Tchoyprakov, J. Fuller
Spatial Integrated Systems, Inc
7524 Standish Place, Suite 100, Rockville, MD 20855
{chang, joe, ali, lelkins} @sisinc.org

ABSTRACT

The problem of tracking the motion of a functional, electrically controlled hand is addressed in this paper. We show that by using skin color and hand segments, the motion of a human hand controlled by a functional electrical stimulation (FES) system can be recovered and predicted. The hand motion tracking and prediction information can then be used as feedback in a closed loop control system to improve the control of the hand motion. We show how skin color tracking can be used to segment the human hand regions and how the hand segments can be used to identify the hand moving positions and how the changes of hand segments between two frames can be used to track the hand orientations. We then show how the hand motion can be recovered based on the changes of the hand positions and orientations and how the motion can be simulated and predicted. The tracking and motion identification methods are verified by the presentation of experimental results obtained using the methods described in the paper.

Keywords hand motion tracking, hand motion prediction, hand position, hand orientation, skin color tracking, hand region, motion reconstruction, recursive least squares

1. INTRODUCTION

Hand motion tracking is a very important area in computer vision applications. The general solutions to this problem are divided into two categories based on the type of sensors used. The two categories are location sensor based hand tracking and vision sensor based hand tracking. In location sensor based hand tracking, sensors such as magnetic tracking equipment are used to identify the 3D hand position. In vision based hand tracking, sensors such as cameras are used to capture 2D views of a moving hand. The latter is widely applied and uses passive sensing, to capture natural motion.

Previous work on vision based hand tracking can be divided into 2D feature-based tracking and 3D model-based tracking. Methods in 2D feature-based register the possible 2D appearances of the moving hand target and find the best matching from the input images [1] [2]. Methods in model-based tracking extract local image features and fit a given 3D shape of hand model to the features [3] [4][5]. B. Stenger [6] used an Unscented Kalman filter to track the hand motion based on a 3D hand model. The model was built from quadrics that approximate the anatomy of a human hand. The filter is used to update its pose in order to minimize the geometric error between the model projection and a video sequence on the background. L. Tsap [7] used a 6 DOF model to simulate motion of a hand and used non-linear Finite Element Method (FEM) to analyze the hand motion from range image sequences. While model-based tracking provides details of finger motion, feature-based tracking deals primarily with hand position. In feature-based tracking, markers, grooves and skin color are used to identify the hand location and algorithms are developed to reconstruct the motion based on the changes of corresponding features over time. Wang [8] devised a set of techniques for segmenting images into coherently moving regions using affine motion analysis and clustering techniques. The scene is divided into four layers, and then the entire sequence is represented with a single image from each layer and with associated motion parameters.

In this paper, we focus on position and orientation tracking of a moving human hand controlled by the FES system.

* This work has been supported by US ONR under grant N00014-02-C-0384.

Optical Pattern Recognition XV, edited by David P. Casasent, Tien-Hsin Chao, Proc. of SPIE Vol. 5437 (SPIE, Bellingham, WA, 2004) · 0277-786X/04/$15 · doi: 10.1117/12.542668

Therefore, feature-based methods are used. FES system uses an electrical activation of the paralyzed neuromuscular system to restore human hand/arm movement and control techniques to individuals with high-level spinal cord injury. In addition, it is often necessary to perform surgical procedures such as muscle tendon transfers [9][10]. The FES system includes sensors, telemetry, stimulators and electrodes. Figure 1 shows an implanted FES Hand Grasp System (HGS).

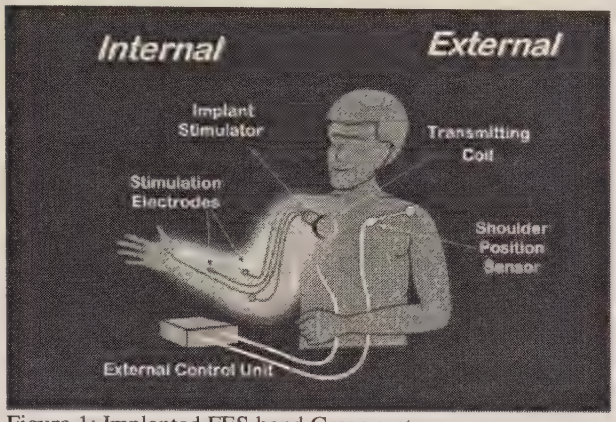


Figure 1: Implanted FES hand Grasp system

The arm motion stimulated by existing HGS is slow, noisy, and has a large amount of error. A position and orientation feedback compensation for the HGS can provide a more precise control of the hand motion. Algorithms developed in this paper are to explore the use of cameras to track the current position and orientation of a moving hand, to predict a position and orientation for the next step, and to provide feedback to the hand motion control system and minimize the deviation from the desired moving trajectory. The algorithm used to update the prediction is based on minimizing the geometric error between the prediction and the observed past positions and orientations from the image frame sequence.

Several simplifying assumptions can be made based on the capabilities of the HGS. Only one hand will be connected to the system and the motion of the hand will be under control of the system. The motion of the hand is reasonably slow and the hand operations are either simply to open or to close the fingers. All motion of the hand, including the operations of the fingers is under control of the HGS. There are two real-time motion tracking cameras are used to track the motion of the hand and there is no tracking of fingers operations.

2. HAND POSITION AND ORIENTATION TRACKING

Hand motion tracking includes identification of moving hand positions and orientations over time. The process of identification is to reconstruct a sequence of 3D points from 2 sequences of images captured from the tracking cameras, the left and right cameras. There are two steps to achieve this goal: first the 2D positions are determined from the captured images, and second, the 2D positions from the two views are used to reconstruct the 3D positions.

The identification of hand moving orientation requires that we reconstruct a sequence of 3D points from the hand positions. A similar two-step procedure can do this, i.e., first, locate the changes of 2D position in current frame from the previous frame, and second, reconstruct the 3D position from 2D positions of two different views.

In order to locate the 2D position of a hand and to find the different 2D positions of the hand segments in the current frame from the previous frame, we need to find the hand segment in the image sequence. Hand region color segmentation is used in this paper. We present hand position and orientation tracking in three sections: 1) Hand regions segmentation by using color image segmentation; 2) hand position tracking; and 3) Hand orientation tracking.

2.1 Hand regions segmentation by using color image segmentation

In this section we discuss hand segment extraction by using color image segmentation. The goal of image segmentation is to classify each pixel in an image into one of a discrete number of color classes called segments. Frequently the number of allowable segments is small; hence, the effect of segmentation is often the identification of the prominent features in an image. Several methods exist – including nearest neighbor and thresholding. In the work described here, a

pixel is thresholded according to its Red, Green, and Blue (RGB) components. The limits and threshold values are found by constructing histograms or RGB values for the entire image [14][15].
In this paper, human skin color segmentation associated with dynamic shareholding method is used to track the hand region in a color image. The basic idea of current color segmentation techniques is to define a color class, and then use it for neighbor classification based on color space thresholding and probabilitistic methods. However, these techniques are susceptible to errors due to lighting. In particular, an incorrect color class may be identified or some critical part of the hand region may be missing.

If two images of the same scene are segmented according to the same threshold values, the various segments form a set of features that can be matched between the two images. For proper feature matching, it is important that the lighting in the images is as nearly the same as possible, and the cameras have similar color settings. Worse results are obtained if color interpretation differs between cameras. A more practical solution for this problem is to use a dynamic shareholding method. In this method, a factor is weighted to a related local color component of a pixel and used to pick a related component. The selection of a pixel is a local issue. Therefore, the lighting in the image can be varying to some degree. In this method, two color classes are used as a set of comparison. The following pseudocode illustrates the process:

$$if((R >= Rlowerthresh)AND(R <= Rupperthresh))$$

$$AND(G >= a * R)AND(G <= b * R)$$

$$AND(B >= c * R)AND(B <= d * R)$$

$$pixel\_color = color\_class;$$

where a, b, c, and d are the weighting factors. In this decision-making algorithm, instead of using fixed threshold values for G and B as in [14], relative comparisons are carried out between G/B and R, thus the impact of lighting conditions is significantly reduced.

$$if((R >= Rlowerthresh)AND(R <= Rupperthresh))$$

$$AND(G >= Glowerthresh)AND(G <= Gupperthresh)$$

$$AND(B >= Blowerthresh)AND(B <= Bupperthresh)$$

$$pixel\_color = color\_class;$$

Figure 2 shows hand color segmentation from a pair of images taken from left camera and right camera.

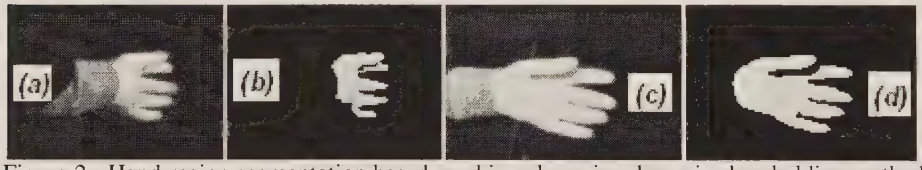


Figure 2: Hand region segmentation based on skin color using dynamic shareholding method. Image (a) and (c) are origin images. Image (a) is with respect to left camera coordinate and (c) to right camera coordinate. Images (b) and (d) are hand segments from (a) and (c).

2.2 Hand position tracking

In this section we discuss 3D hand position extraction by using 2D hand segments from the tracking cameras. The goal of hand position extraction is to reconstruct the location of a hand in a 3D space.

In this paper, we define the position of a hand be the center of a hand in 3D space and assume the projection of the hand position falls into the center of a hand segment. The 3D hand position with respect to the robot coordinates in the hand motion simulations system can then be calculated. Figure 3 illustrates how the 3D hand position is constructed from two hand segments:

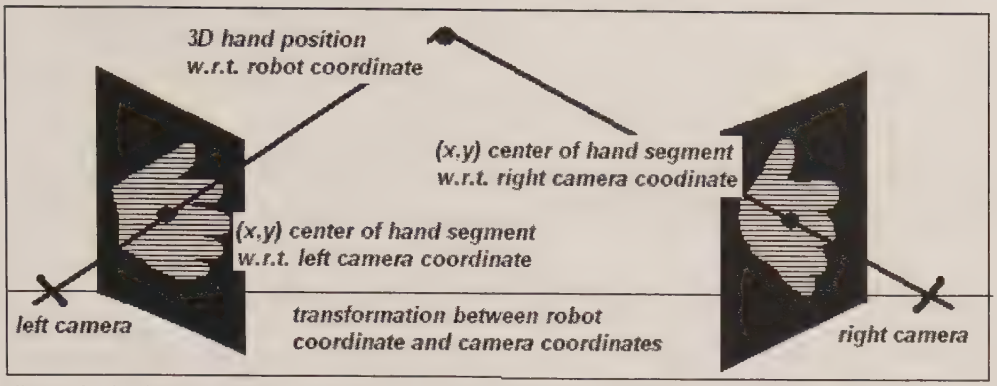


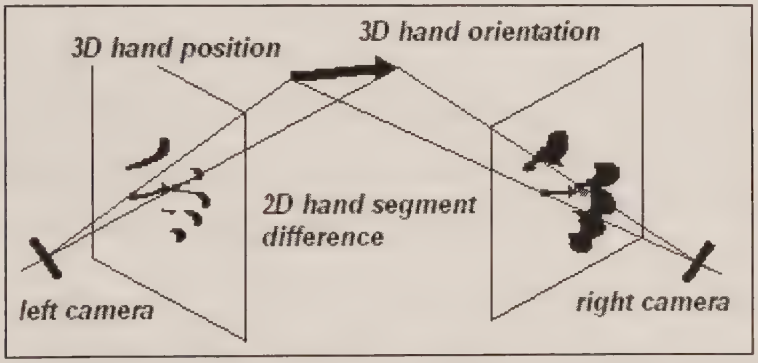
Figure 3: 3D hand position with respect to robot coordinate reconstruction from hand segments of left camera and right camera.

The locations of two tracking cameras are determined during system calibration. Then the transformations from left and right camera coordinates to global coordinates (in this paper, global coordinate is referred to robot coordinate) are used to transfer the center point of the hand segment from two tracking cameras to the robot coordinates. By finding the intersection of the two straight lines determined by the center point and the cameras, the 3D hand position is then calculated.

2.3 Hand orientation tracking

In this section we discuss 3D hand orientation extraction by using difference of 2D hand segment between the current frame and the previous frame. The goal of hand orientation extraction is to reconstruct the moving direction of the hand over time.

In this paper, we define the orientation of a hand by assigning the Z- axis to the middle finger direction and the Y-axis normal to and pointing away from the palm of the hand as shown in right park of Figure 4.



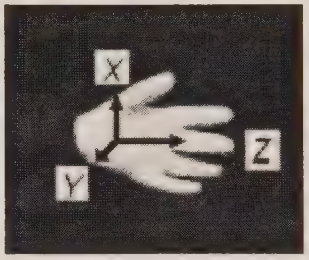


Figure 4: Hand orientation assignment and tracking from difference images between hand segments from frame to frame.

Hand orientation tracking is illustrated in the left hand part of Figure 4. When a hand is moving over time, the difference between hand segments from frame to frame can be calculated. The regions of the difference indicate the 2D projection of a change in 3D space. It can be seen from the Figure that a hand has moved from left to right and rotated a positive angle about the Z-axis (assume that the Z-axis of the robot coordinate is up). This is evident from the motion vector. It can be seen from the figure that in this example, changes observed from the right camera is larger than the changes observed from the left camera. In this algorithm, the hand orientation tracking is calculated following steps:

a) At a time t, take images $I_{(L,t)}$ from left camera and $I_{(R,t)}$ from right camera.

b) Calculate hand segments $S_{(L,t)}$ from $I_{(L,t)}$ and $S_{(R,t)}$ from $R_{(R,t)}$.

c) Reconstruct hand 3D position P_t using methods discussed in **2.1**.

d) Calculate hand segment difference image $S_{(L,d)}$ by subtracting $S_{(L,t-1)}$ from $S_{(L,t)}$ and $S_{(R,d)}$ by subtracting $S_{(R,t-1)}$ from $S_{(R,t)}$.

e) Calculate 2D center points $C_L(x, y)$ from $S_{(L,d)}$ and $C_R(x, y)$ from $S_{(R,d)}$.

f) Reconstruct hand orientation 3D position Q_t from $C_L(x, y)$ and $C_R(x, y)$ using methods discussed in **2.1**.

g) The orientation of the hand orientation O_t at time t is the 3D vector connected P_t and Q_t.

3. HAND MOTION MODLE PARAMETERS ESTIMATION

In order to recover the motion of a hand, it is necessary to construct a hand motion dynamic mode. The hand motion dynamic mode is a mathematical model describing the changes of the hand position state from time to time. How accurate the hand motion dynamic mode can describe the motion of a hand is based on two factors: how suitable the dynamic characteristic mode is created and how precisely the parameters of the mode are selected. The discussion in creation of hand motion dynamic mode and selection of mode parameters is divided into two subsections: 1) hand motion model and its parameters; 2) hand motion model parameter estimation.

3.1 Hand motion model and its parameters

If we focus at its positions rather than its gestures, the hand motion can be described by the motion of a point in 3D space The motion of a 3D point can be represented as changes of state of a 3D point. The changes of a state can be described by a dynamic system. As a dynamic system, the motion can be described by the evolution of a set of state variables that consists of three rotation parameters and three translation parameters. Let $X=\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}'$ be a point on the curve that describes the motion of a point. Subsequent points on the curve may be obtained by rotating X by a small angle α about some axis, says A, and then by translating the rotated point by the translation vector $T=\begin{bmatrix} t_x & t_y & t_z \end{bmatrix}'$. The three translation parameters referred to above are t_x, t_y, and t_z.

To determine the three rotation parameters, the Euler angles, α_x, α_y, and α_z, are calculated for the axis A. Rotation about A by α is then equivalent to rotating about the Z-axis by α_z, then rotating about the Y-axis by α_y, and then rotating about the X-axis by α_x. The rotation matrix about the axis A is the product of the 3 rotation matrices about the coordinate axes. Hence the model for rotation followed by translation can be written as:

$$X_{t+1}=RX_t+T$$

Where

$$R=\begin{bmatrix} \varpi_0 & \varpi_1 & \varpi_2 \\ -\varpi_1 & \varpi_0 & \varpi_3 \\ -\varpi_2 & -\varpi_3 & \varpi_0 \end{bmatrix} \qquad (2)$$

The values ϖ_1, ϖ_2 and ϖ_3 from the rotation matrix R are the three rotation parameters described above and ϖ_0 is assumed to be 0 - along with t_x, t_y, and t_z form the set of six parameters [11].

To model the motion of a hand with noise, we add a noise vector to equation (1). Define the rotation and translation noise vector $e=\begin{bmatrix} e_1 & e_2 & e_3 \end{bmatrix}$, where e_1, e_2, e_3 are independent normally distributed random noise. Since the noise is additive, a transformation matrix c with diagonal entries c_1, c_2 and c_3 may be defined. Then equation (1) becomes:

$$X_{t+1}=RX_t+T+ce \qquad (3)$$

$$c=\begin{bmatrix} c_1 & 0 & 0 \\ 0 & c_2 & 0 \\ 0 & 0 & c_3 \end{bmatrix} \quad (4)$$

3.2 Hand motion parameter estimation

In section **3.1**, a dynamic mode of hand motion has been developed and a discrete state equation was obtained as shown in Equation (3). In this section we discuss how the parameters of the state equation, the rotation matrix R and translation vector T can be determined.

If we rewrite Equation (3) so that the parameters from rotation matrix R and translation vector T are separated from the state parameters which can be observed as previous discussion in section **2**. Obviously, Equation (3) can be written in the same format as a linear regression model of

$$X_{k+1}=\phi'_k\theta \quad (5)$$

Where

$$X=\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \quad (6)$$

$$\phi'=\begin{bmatrix} x_1 & x_2 & x_3 & 0 & 1 & 0 & 0 & e_1 & 0 & 0 \\ x_2 & -x_1 & 0 & x_3 & 0 & 1 & 0 & 0 & e_2 & 0 \\ x_3 & 0 & -x_1 & -x_2 & 0 & 0 & 1 & 0 & 0 & e_3 \end{bmatrix} \quad (7)$$

$$\theta'=\begin{bmatrix} 1 & \varpi_1 & \varpi_2 & \varpi_3 & t_x & t_y & t_z & c_1 & c_2 & c_3 \end{bmatrix} \quad (8)$$

Equation (7) is a parameter vector included six transformation and three noise parameters. The problem of transformation parameters estimation now becomes a problem of constructing a general mapping from a sequence of observation data, starting from the observed data, recursive methods pose further constraints on the computation of parameters estimates. It is easy to consider that by taking a guess and then modifying it. Thus, equations (5) can be easily solved by applying a recursive least squares algorithm [12] [13]:

$$\hat{\theta}_t=\hat{\theta}_{t-1}+K_t[y_t-\phi_t\hat{\theta}_{t-1}] \quad (9)$$

$$P_t=P_{t-1}+P_{t-1}\phi_t[I+\phi'_tP_{t-1}\phi_t]^{-1}\phi'_{t-1}P_{t-2} \quad (10)$$

Where

$$K_t=P_{t-1}\phi_t[I+\phi'_tP_{t-1}\phi_t]^{-1} \quad (11)$$

With initial condition of P_0 positive $\hat{\theta}_0$ are given.

4. EXPERIMENTAL RESULTS

The algorithms described above were implemented and tested in an actual setting. We present the results in three sections: 1) hand motion tracking simulations system configuration; 2) hand position and orientation tracking; 3) hand motion recovery and prediction. Implementation of the algorithms and tracking simulation was done using a robot system. A human hand model was attached to the end-effector of the robot.

4.1 Hand motion tracking simulations system configuration

In this section, we demonstrate how a hand motion simulation system is set up and how the simulated hand motion is observed. For the experimental purpose, a model of a human hand is used to simulate a real human hand and a 6-DOF robot is used to simulate motion of a human arm as shown in Figure 5.

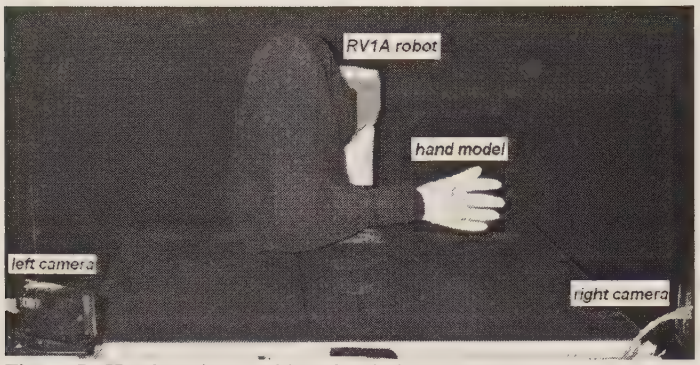


Figure 5: Hand motion tracking simulation system

In this experiment, a model of a human hand is attached to the end-effector of a Mitsubishi RV1A robot. Two tracking cameras, the "left camera" and the "right camera" are located in front of the moving hand model. For the purpose of speeding up color image segmentation during the experiment, the robot body and the working background were covered with dark clothes. The tracking cameras capture a sequence of images while the robot moves the hand model. Two set of images with their corresponding frame number from the 'Left camera" and the "right camera" are shown in Figure 6.

Figure 6: Images taken from left and right cameras with frames numbers

From the sequence of the image frames one can see different outlooks of the hand based on two different views of two tracking cameras. Numbers of the frames from the sequence are picked in order to see the obvious changes of the view.

4.2 Hand position and orientation tracking

In this section, we demonstrate how closely the hand positions and orientations are tracked. When the image frames captured as shown in Figure 6 were segmented and the centers of the segments were found, a sequence of 3D points corresponded to the centers were calculated. Figure 7 shows the hand position-tracking trajectory compared with the real hand motion trajectory.

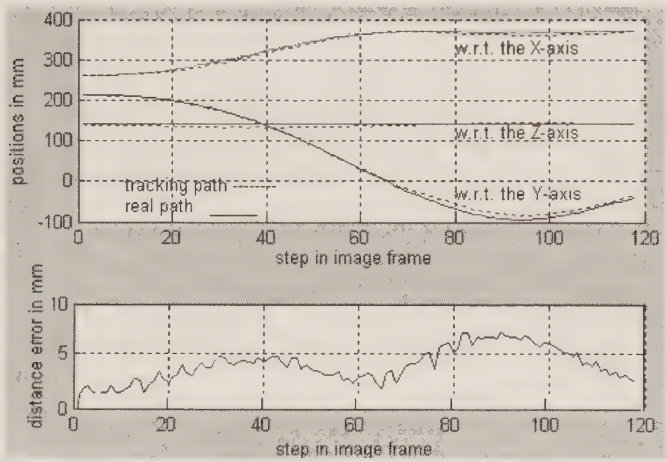


Figure 7: Hand position tracking trajectory compared with real motion trajectory.

From the figure, one can see while the tracking positions follow the real motion of the hand model. A periodical error appears corresponding to the motion of hand with respect to the X-axis, the Y-axis and the Z-axis. This is because of the calculation errors of the hand position. When the hand moves, its orientation to the tracking cameras are changing. The change causes the change in calculation of the hand segments. When a hand segment is large enough, the calculation of the hand center from the segment is much more accurate. On the other hand, as the hand segment becomes smaller, the error to extract the center can be larger.

Errors in 3D distance between tracking trajectory and real motion trajectory appear with periodical characteristics as shown in second part of Figure 7. Similar to its periodical characteristics with respect to different axes, the position tracking errors in 3D distance have its periodical characteristics as shown in Figure 8. This in fact provides an opportunity to design a filter to correct the position tracking based on change of hand orientations.

While the absolute hand orientation tracking errors are still smaller than 1 degree, the error distribution is irregular as shown in Figure 8.

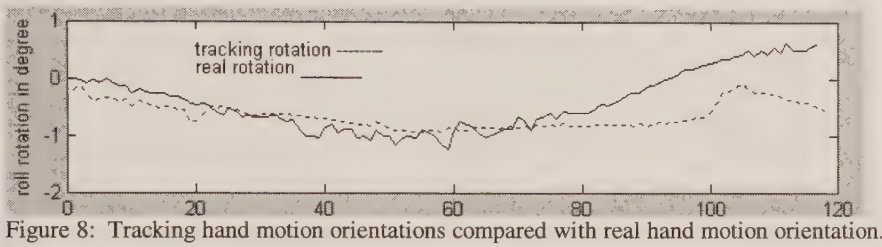


Figure 8: Tracking hand motion orientations compared with real hand motion orientation.

4.3 Hand motion recovery and prediction

In this section, we demonstrate the ability of RLS algorithm to perform hand motion parameter estimation. A sequence of hand motion 3D points, shown in Figure 7, is used as the observed values for equation (7). Values as large as 10^{12} are used for the initial values of the covariance matrix P_0. The initial estimate parameter vector $\hat{\theta}_0$ is set to $[1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0]^t$. The initial noise matrix is initialized randomly by $c_{(1,1)} = 0.5$, $c_{(2,2)} = -0.6$ and $c_{(3,3)} = 0.5$. A set of 40 unit random white noisy points together with the parameter vector of equation (4) and the state matrix of equation (3) were used. The RLS algorithm is allowed to continue through 40 iterations. The results are shown in Figure 9.

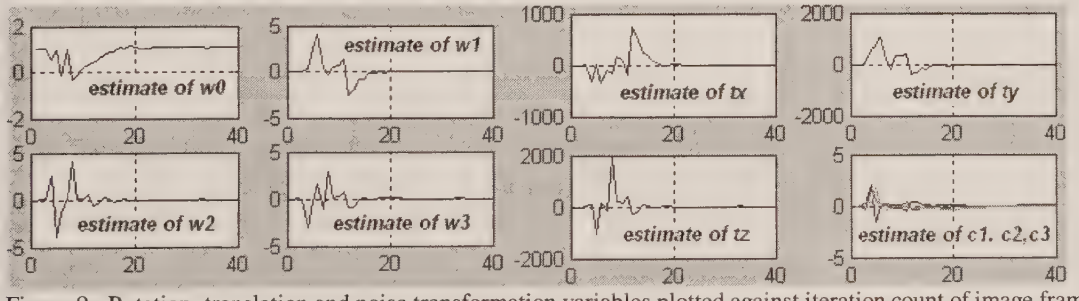


Figure 9: Rotation, translation and noise transformation variables plotted against iteration count of image frame number for the RLS algorithm.

The estimates of the parameter vector reach a stable value after about 20 iterations. A set of the computed estimates is obtained and the parameter estimations are plotted in Figure 10. The final computed estimation parameters are:
Estimated rotation variables: 0.0053 -0.0063 -0.0008
Estimated translation variables: 4.3559 0.0528 -0.9922
Estimated noise variables: -0.0069 -0.0878 -0.0369

A sequence of prediction of the hand motion was calculated based on the estimate parameters and compared with the tracking hand positions as shown in the first part of Figure 10. Detailed plots of comparisons between the prediction and tracking positions with respect to different axes are also shown in second part of Figure 10.

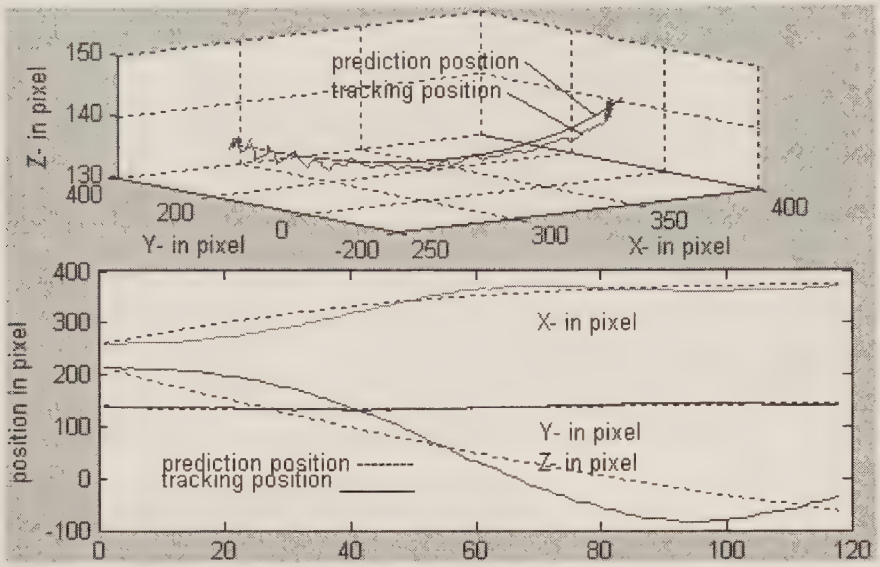


Figure 10: Hand position prediction and tracking comparison in 3D space and details comparison with respect to all axes.

5. CONCLUSION AND FUTURE WORK

In this paper, the motion recovery of a FES controlled hand is proposed. In this method, limited hand motion without finger operations can be tracked and motion parameter estimation is conducted by hand region segmentation and minimizing the errors between the predicted results and the observed results. The effectiveness of the proposed method has been demonstrated through experiments using real images. Future work will be hand motion recovery with finger operations based on this work.

ACKNOWLEDGEMENT

The authors wish to thank H. Deng for valuable discussion in edge detection issues and his helpful comments.

REFERENCES

[1] B.Moghaddam and A. Pentland, "Maximum likelihood detection of faces and hands", *Proc. Of Int. Workshop on Automatic Face and Gesture Recognition*, pp. 122-128, June 1995.
[2] U. Brockl-Fox: "Real-Time 3-D Interaction with up to 16 Degrees of Freedom from Monocular Video Image Flows ", *International Workshop on Automatic Face and Gesture-Recognition*, 1995, University of Zurich, pp. 172- 178.
[3] J. Rehg and T. Kanade, "Visual Tracking of High DOF Articulated Structures: an Application to Human Hand Tracking", ECCVV'1994, pp.35-46, 1994.
[4] David G. Lowe, "Fitting Parameterized Three-Dimensional Models to Images", *IEEE Trans. On PAMI*, Volt 13, No 5, pp 441-450, 1991.
[5] M. Mochimaru and N. Yamazaki. "The Three-dimensional Measurement of Unconstrained Motion Using a Model-matching Method". *ERGONOMICS, vol.37, No.3*, pp. 493–510, 1994.
[6] B. Stenger, P. R. S. Mendonça, and R. Cipolla, "Model-based hand tracking using an unscented kalman filter," in *Proc. British Machine Vision Conference*, vol. I, (Manchester, UK), pp. 63–72, September 2001.
[7] L. Tsap, et al., "Human skin and Hand Motion Analysis from Range Image Sequences Using Nonlinear FEM", *IEEE Proc. Nonrigid and Articulated Motion Workshop*, pp. 80-88, 1997.
[8] John Y. A. Wang and Edward H. Adelson "Layered Representation for Motion Analysis", *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 361-366, New York; June 1993.

[9] Perreault, E.J., Kirsch, R.F., and Acosta, A.M. "Multi-input, multi-output system identification for characterization of limb stiffness dynamics", *Biol. Cybern.* 80:327-337, 1999.
[10] Abbas JJ, Triolo RJ. "Experimental Evaluation of an Adaptive Feed forward Controller for Use in Functional Neuromuscular Stimulation Systems", *IEEE Transactions on Rehabilitation Engineering*, 5(1): 12-22, 1997.
[11] T.S. Huang, "Image sequence analysis", Spring-Verlag Berlin Heidelberg, 1981.
[12] S. Chang, "Rigid object motion parameter estimation from a dynamic image sequence", Ph.D. dissertation, George Mason University, Fairfax, VA, 1997.
[13] S. Chang, J. Fuller, A. Farsaie, L. Elkins, "Recursive least squares approach to calculate motion parameters for a moving camera", *SPIE International Symposium on Photonics Technologies for Robotics, Automation, and Manufacturing*, October, 2003.
[14] J. Bruce, T. Balch, M. Veloso, "Fast and Cheap Color Image Segmentation for Interactive Robots", School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213.
[15] S. Chang, J. Fuller, A. Farsaie, L. Elkins, "Color Feature and Density Based Image Mosaicing Using Repeated Application of the ICP Algorithm', *IS&T/SPIE 16th Annual Symposium on Electronic Imaging Science and Technology*, January, 2004.

Enhanced Projection Slice Theorem Synthetic Discriminant Functions Based on the Karhunen-Loeve Transform with Application to the Protein Structure Identification in Cryo-Electron Microscopic Images

Vahid R. Riasati, and Hui Zhou
California State Polytechnic University-Pomona
Electrical and Computer Engineering
3801 West Temple Avenue
Pomona, CA 91768

Abstract

In this paper we utilize the Karhunen-Love Transform with the Projection-Slice Synthetic Discriminant Function Filters, KLPSDF to reduce the data set that represents each of the training images and to emphasize the subtle differences in each of the training images. These differences are encoded into the PSDF in order to improve the filter sensitivity to the recognition and identification of protein images formed from a cryo-electron microscopic imaging process. The PSDF has been shown to improve the performance of the SDF on specific images in previous papers. The protein structures found in cryo-electron microscopic imaging represent a class of objects that have low resolution and contrast and subtle variation. This poses a challenge in design of filters to recognize these structures due to false targets that often have similar characteristics as the protein structures. The incorporation of the KLT in forming the filter provides an optimal method of decorrelating images prior to their incorporation into the filter. We present our method of filter synthesis and the results of the application of this modified filter to a protein structure recognition problem.

Key words/phrases: Projection-Slice Theorem, PSDF, Synthetic Discriminant Functions, SDF, Optical Pattern Recognition, Karhunen Love Transform, Principle Components.

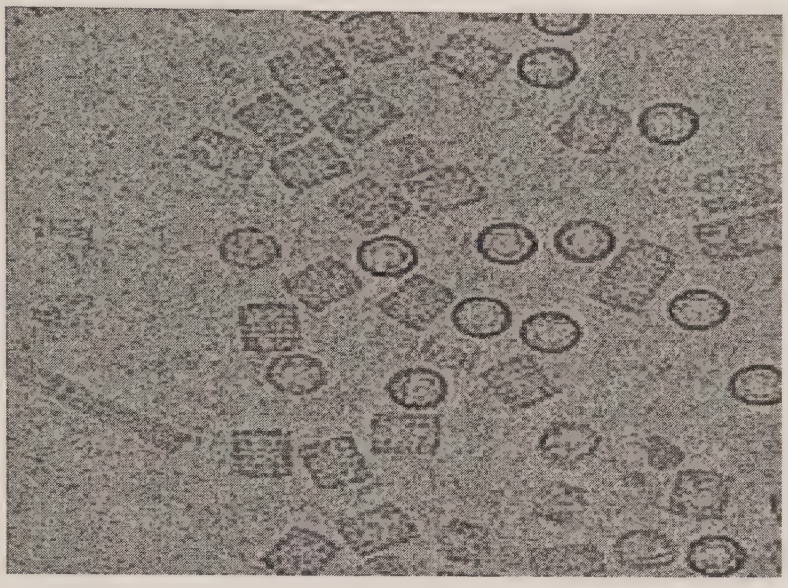
Background

The Karhounen Loeve Transform Projection-Slice Synthetic Discriminant Function, KLPSDF, filter is a variation of the Projection-Slice Synthetic Discriminant Function, PSDF, filter[1,2]. The PSDF filter involves the incorporation of complete training data in the filter. The large unaltered amounts of data representing each training pattern leads to data overlap and nulls the matching conditions upon which SDF filters were first conceived [3]. To reduce the data overlap problem, the PSDF utilizes a set of projections to represent each of the training images[5,6]. The appropriately ordered projections, refer to different illumination directions for each target and the combined projections from each training image reduces the data and improves the filter performance. In this work, we use the Karhounen Loeve Transform, KLT, [7] in the PSDF synthesis process. In two-dimensions, 2-D, the KLT effectively seeks to re-align the training images near the least variant directions, largest Eigen-vectors, of the covariance matrix of the combination of training images. This utilization of the orthogonal directions of each of the training images allows for optimal representation of the training images as a set of matrices. Hence, this set of training images provides the relative non-redundant variation of our training images. The KLT resultant non-redundant data set represents the training images as a set where each image relates to the other images in the set and provides information that is not provided by the other images. Each of these new training images are represented by a set of projections, and the projections are combined via the usual PSDF algorithm. Though, this technique is general and applicable to any set of training images, we have decided to apply our technique to the problem of protein structure recognition.

The computational problem is the identification of each of the 2-D projections of three-dimensional structures. The isolation and removal of the 2-D projections from the larger conglomerate image is a secondary issue. We selected eight training images in two sets represented on a single cryogenic-

Optical Pattern Recognition XV, edited by David P. Casasent, Tien-Hsin Chao, Proc. of SPIE Vol. 5437 (SPIE, Bellingham, WA, 2004) · 0277-786X/04/$15 · doi: 10.1117/12.547154

electron recording of 2-D projections of protein structures to implement our SDF, and KLTSDF filters. The original image [8]containing the of 2-D projections of the protein structures is shown in Figure (1).



Figure(1), The original electron microscopic image of the 2-D projections of the protein structure

In an ideal scenario the filter is synthesized with the minimum amount of information and performs perfect matching of all variations of the recorded 2-D projections of the protein structures. In our scenario, four training images were chosen in different sub-image sizes. These training images are shown in Figure (2), below,

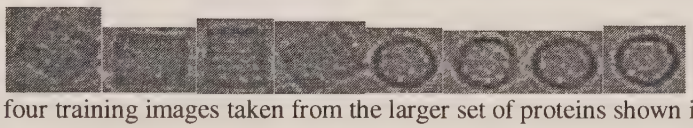


Figure 2. two sets of four training images taken from the larger set of proteins shown in Figure (1).

This angle is equivalent to the irradiating beam angle that produces the projection. There has been a significant effort in the utilization of the PST for the reconstruction of images [6,9]. It has been shown that an image can be represented by a discrete set of 1-D projections taken at different projection angles. The utilization of this approach in the implementation of the SDF filter has shown to improve the recognition capabilities of the filter, and reduce the data requirements for the synthesis of filters [1, 2, 10]. Instead of using the entire image in the synthesis of the SDF filter we used only a finite number of projections to represent each of the training images. It was shown that the PST could be utilized to reduce the normal two-dimensional SDF problem of combining matched filters to a problem of combining several 1-D slices of each of those training images in the Fourier domain (FD). The spatial separation of each of the slices used in the PSDF limits the number of training images that can be represented in a finite matrix of points that represent the PSDF filter. Previously, we looked at the utilization of techniques to reduce the data in an attempt to incorporate more pertinent and feature-related data in the PSDF filter. This was attained via the wavelet transform [11]. The key factor utilized was to attain comparable, or even improved, performance as the performance of the PSDF after data were removed from the filter. In this work, we look at improving the performance of the PSDF filter by being choosy about the slices that we select to implement the PSDF filter. To do this, we have chosen The wavelet transforms have proven to be a .sophisticated tool for image compression through their ability to identify key features of patterns. The hope is to be able to achieve a similar filter performance after the compression as that obtained prior to the compression of the projections of the objects.

Theory and Discussion:

The KLT is first utilized to generate the new training training images. This is done by calculating the mean for each training image, as

$$\mathbf{m}_S = \frac{1}{M}\sum_{k=1}^{M} S_k \tag{1}$$

Where, S, represents the lexicographical ordering of the training images, and

$$\mathbf{m}_S = E[S], \tag{2}$$

is the mean of S and is approximated by Eq. (1)., using the mean to find the covariance matrix, given by,

$$\mathbf{C}_{S-m_S} = E\{(S-\mathbf{m}_S)(S-\mathbf{m}_S)^T\}, \tag{3}$$

and approximated by,

$$\mathbf{C}_{S-m_S} = \frac{1}{M}\sum_{k=1}^{M} S_k S_k^T - \mathbf{m}_S\mathbf{m}_S^T \tag{4}$$

the creation of the new training images can be completed by re-aligning the resulting covariance along a direction where the matrix is a constant multiplier, ie. calculate the eigen values and re-order the image relative to the largest eigen values as each of these training images represent the most that is in common in the training imaes, or the least variation relative to the training images. We use this information to classify our objects. The example, would be to use anything with two up-right legs and two arms and a head as a human.

Because $\mathbf{C}_{S-m_S}$ is real and symmetric, finding a set of n orthonormal eigenvectors always is possible. Let e_i and λ_i, $i = 1,2,\cdots\cdots,n$, be the eigenvectors and corresponding eigenvalues of $\mathbf{C}_{S-m_S}$, arranged (for convenience) in descending order so that $\lambda_j \geq \lambda_{j+1}$ for $j = 1,2,\cdots\cdots,n-1$. The KLT effectively seeks to re-align the training images near the least variant directions, largest Eigen-vectors, of the covariance matrix of the combination of training images. This utilization of the orthogonal directions of each of the training images allows for optimal representation of the training images as a set of matrices. Hence, this set of training images provides the relative variation of our training images. If **A** be a matrix whose rows are formed from the eigenvectors of C_x, ordered so that the first row of **A** is the eigenvector corresponding to the largest eigenvalue and the last row is the eigenvector corresponding to the smallest eigenvalue. Suppose that A is a transformation matrix that maps the **x**'s into vectors denoted by **y**'s, as follows:

$$\mathbf{y} = \mathbf{A}(S-\mathbf{m}_S) \tag{5}$$

This is called the Karhunen-Loeve Transform. The mean of the y vectors resulting from this transformation is zero; that is $\mathbf{m}_\mathbf{y} = 0$. And the covariance matrix of the **y**'s can be obtained in terms of **A** and $\mathbf{C}_{S-m_S}$ by

$$\mathbf{C}_\mathbf{y} = \mathbf{A}\mathbf{C}_{S-m_S}\mathbf{A}^T \tag{6}$$

$\mathbf{C}_{S-m_S}$ and C_y also share same eigenvalues and eigenvectors.

Another important property of the Karhunen-Loeve Transform deals with the reconstruction of S from **y**. Because the rows of A are orthonormal vectors, $\mathbf{A}^{-1} = \mathbf{A}^T$, and any vector x can be recovered from its corresponding y by using the relation:

$$S = \mathbf{A}^T\mathbf{y} + \mathbf{m}_S \tag{7}$$

Suppose, however, that instead of using all the eigenvectors of $\mathbf{C}_{S-m_S}$ we form matrix $\mathbf{A}_k$ from the K eigenvectors corresponding to the K largest eigenvalues, yielding a transformation matrix of order $K \times n$. The y vectors would then be K dimensional, and the reconstruction given in equation (6) would no longer be exact. The vector reconstructed by using $\mathbf{A}_k$ is

$$\hat{S} = \mathbf{A}_k^T\mathbf{y} + \mathbf{m}_S \tag{8}$$

It can be shown that the mean square error between S and $\hat{S}$ is given by the expression

$$e_{ms} = \sum_{j=1}^{n} \lambda_j - \sum_{j=1}^{k} \lambda_j = \sum_{j=k+1}^{n} \lambda_j \qquad (9)$$

The first part of Equation 8 indicates that the error is zero if $K=n$ (that is, if all the eigenvectors are used in the transformation). Because the λ_j's decrease monotonically, Equation (8) also shows that the error can be minimized by selecting the K eigenvectors associated with the largest eigenvalues. Thus the KL transform is optimal in the sense that it minimizes the mean square error between the vectors S and their approximations $\hat{S}$.

The PSDF utilizes 1-D Fourier slices from different distortions, or training images, to form a 2-D image for the recognition of the object, or class of objects, after it/they have undergone a distortion, even if the new distorted image is not a training image. If M equi-spaced slices are taken from the FT of each of N-training images, the separation in angle between these M-slices can be given by $\frac{\pi}{M}$. This implies that if the nth training image has slices at

$$\theta_{mn} = \frac{m\pi}{M} + \frac{n\pi}{MN} \qquad \text{for } n = 0,1,\ldots,N-1, \text{ and } m = 0,1,\ldots,M-1. \qquad (10)$$

where θ_{mn}, is the angle for the mth slice of the nth training image, then the slices would not, in theory, overlap, except at zero frequency. Utilizing the generalized PSDF filter [2] transfer function with the training image estimates given in Eqn. (4.8) the new KLPSDF can be defined as

$$H(u,v) = \sum_{n=0}^{N-1}\sum_{m=0}^{M-1} a_n \hat{S}_n(\omega\cos\theta, \omega\sin\theta)\delta(\theta - \theta_{mn})|\omega|^{\gamma} \qquad (11)$$

where, the coefficients, a_n, are found by following standard SDF design procedure [3], $\hat{S}_n$ is the Fourier transform of the nth training image estimate, θ_{mn} is the projection angle, and m and n are as defined in Equation (1). The frequency domain has been described in the continuous polar coordinates, (ω, θ), $\delta(\bullet)$ is the Dirac delta function, and the $|\omega|$ is a result of calculating the Jacobean [6] necessary for the change from rectangular to polar coordinates [1,2,9]. We have used γ as a real number greater than zero in the filter transfer function in order to improve the performance of the filter[2]. Taking an inverse FT of equation (2) results in the filter impulse response that is given by,

$$h(x,y) = \frac{1}{2\pi}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1} a_n \hat{g}_{\theta_{mn}}(t), \qquad (12)$$

where, $\hat{g}_{\theta_{mn}}(t)$, is given by

$$\hat{g}_{\theta_{mn}}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{\theta_{mn}}(\omega)e^{j\omega t}\,|\omega|^{\gamma+1}\,d\omega \qquad (13)$$

$$= p_{\theta_{mn}}(t) * \hat{k}(t)$$

where, * denotes convolution, the independent variable, t, is given by

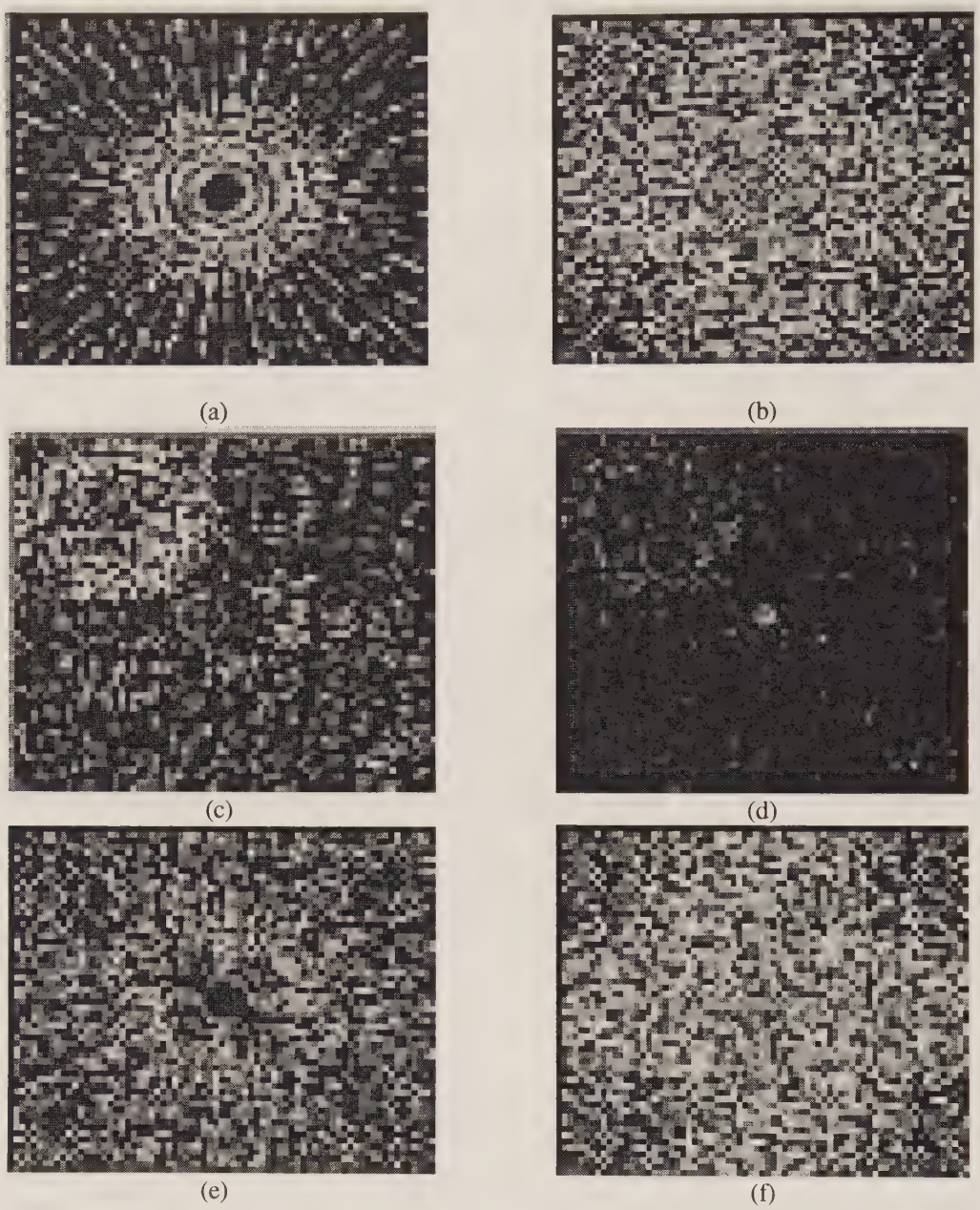
$$t = x\cos\theta_{mn} + y\sin\theta_{mn}, \qquad (14)$$

$p_{\theta_{mn}}(t)$ is the inverse FT of the Fourier slice, $\hat{S}_{\theta_{mn}}(\omega)$, taken at θ_{mn}, and the function $\hat{k}(t)$ is the inverse FT of $|\omega|^{\gamma+1}$, which exists in generalized form [2, 6].

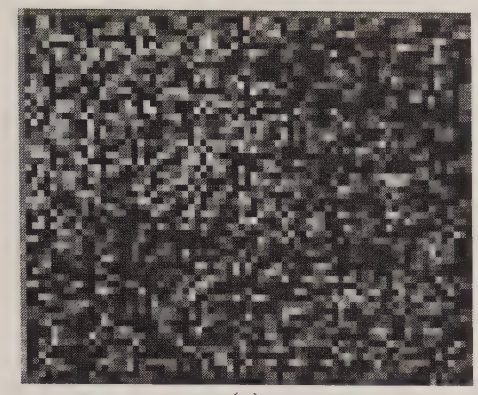
Simulation Results

The PSDF filter has been implemented by extracting 20-slices from each of the training images. These slices are combined in order, as given in Equation (2), with a γ=1.3. The sum of the weighted slices from each of the training images represent the PSDF filter transfer function shown in Figure (4).

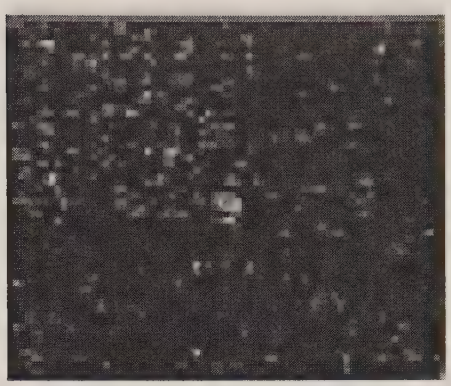
Using the set of training patterns shown in figure (2), the PSDF filter is generated in complex and phase-only, PO, form. Figure (4) show the impulse response and the transfer functions of the complex and the PO implementation of the SDF, respectively.



(a) (b) (c) (d) (e) (f)



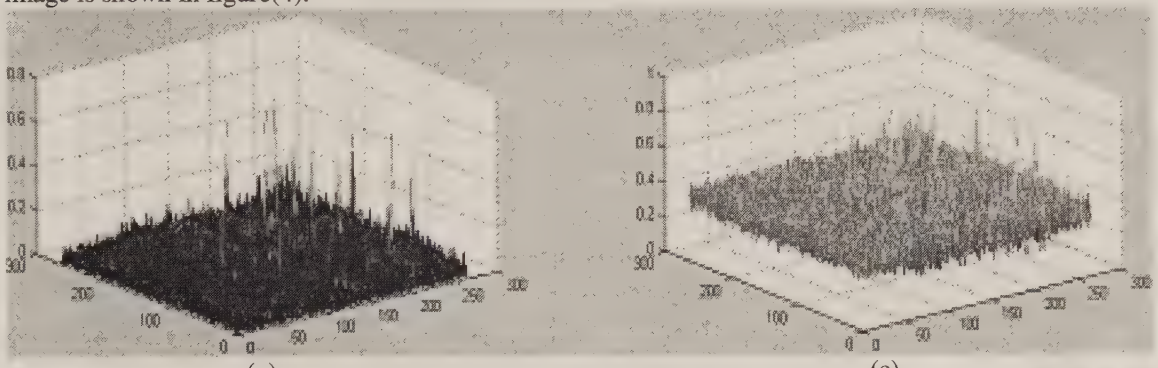
(g)



(h)

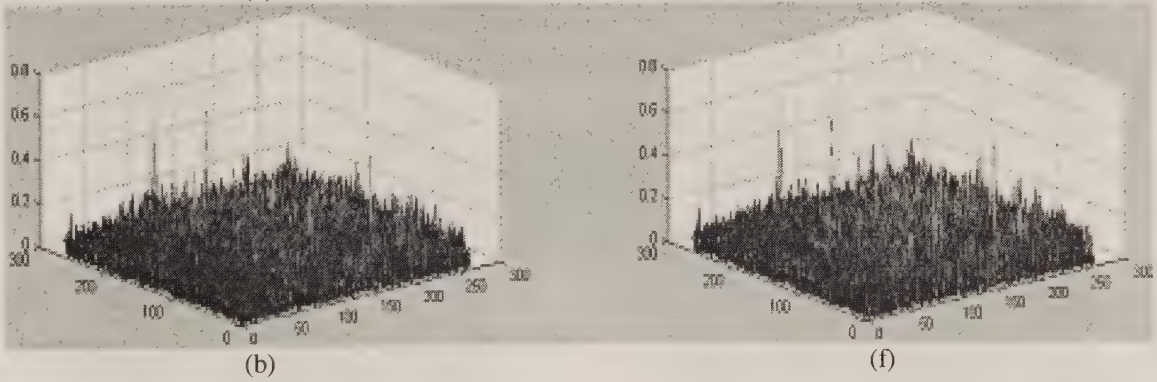
Figure(3) The top/bottom and side-view PSDF filter (a)&(e) transfer function magnitude, (b)&(f) transfer function phase, (c)&(g) impulse response, (d)&(h) and POF impulse response, shown for each of the sets of 4-training images, respectively, each of the side view, and bottom/top view identification of the protein structures.

The results of the application, by means of correlation, of the SDF filter generated above to the original image is shown in figure(4).



(a)

(e)



(b)

(f)

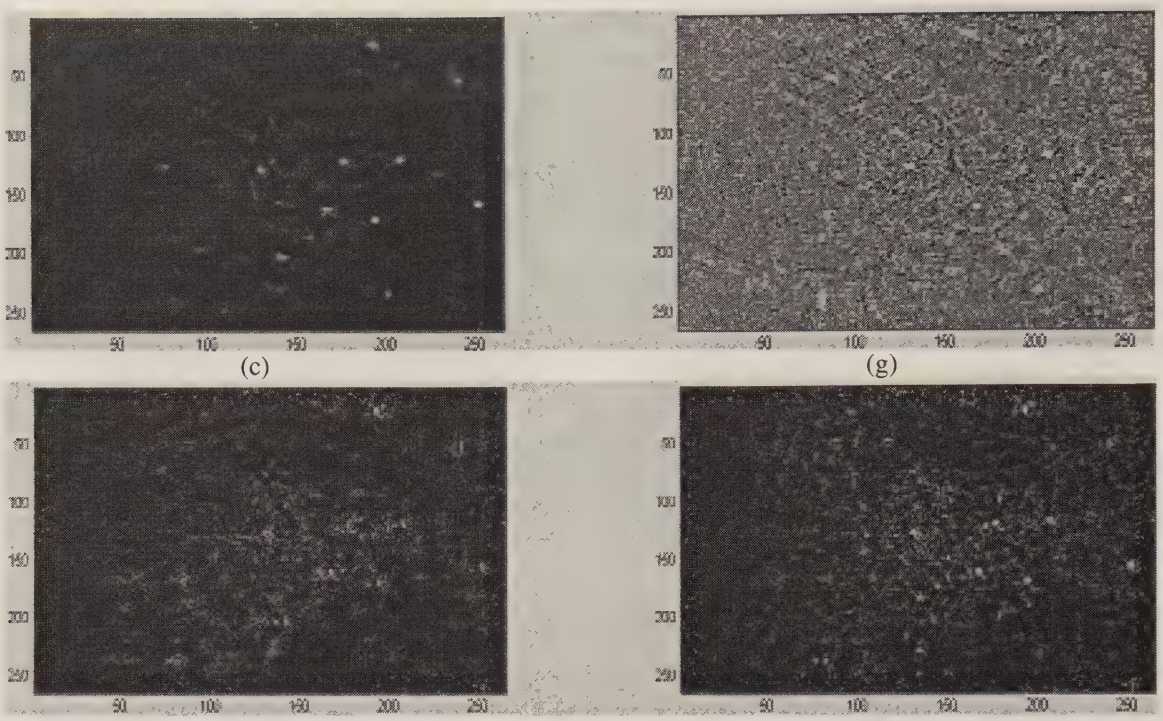


Figure (4) The correlation results of the original image with the PSDF filter in (a) through (d) show the complex implementation for the bottom/top and side views, respectively, and (e), and (h) show the phase-only implementation the bottom/top and side views, respectively.

For the protein structures, we used the same four training images that we used with the PSDF. The KLPSDF filter using each of the two sets of four components shown in Figure(5), and combined these components. The first four principle components are shown in Figure 5.

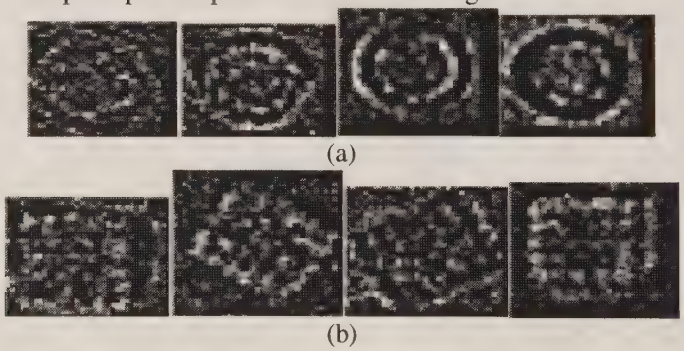
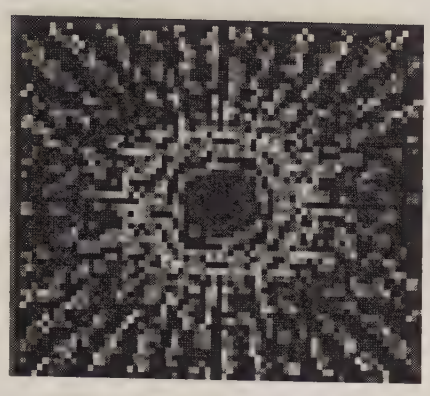


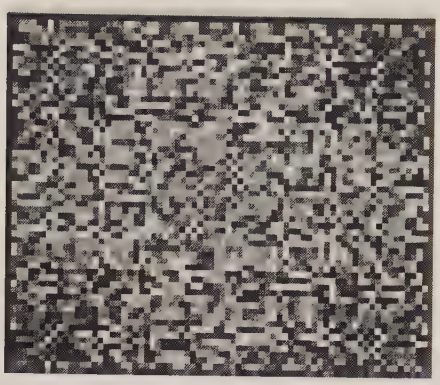
Figure (5) The four principle components of used for training each of the two KLPSDF filters.

In order to increase the identification of the class of structures for the PSDF, the KLPSDF has been implemented, as given in Equation (6), for the same training images, the same γ=1.3, and for 20-slices from each of the training images. This new filter is shown below, in Figure(6).

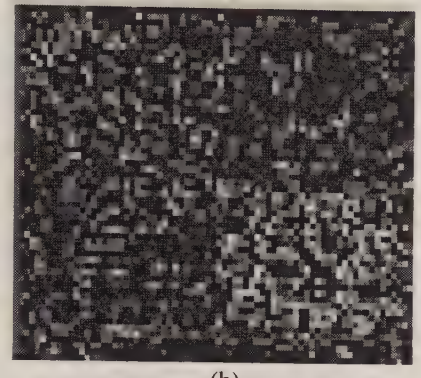
Each of the two KLPSDF filters were implemented in complex and phase-only form. The transfer function and impulse response of these implementations are shown in Figure(6).



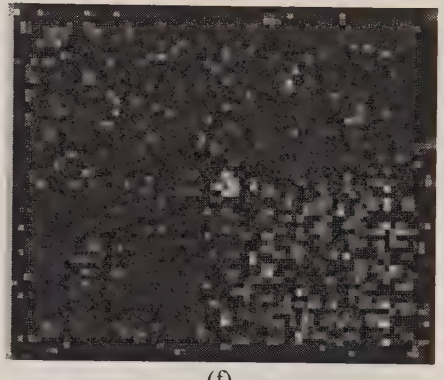
(a)



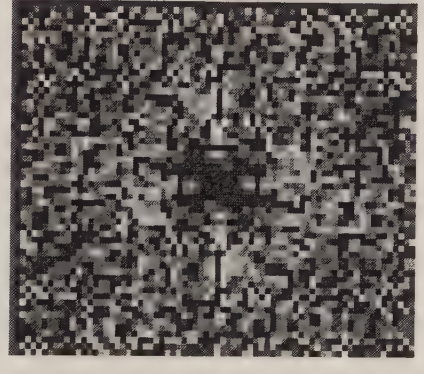
(e)



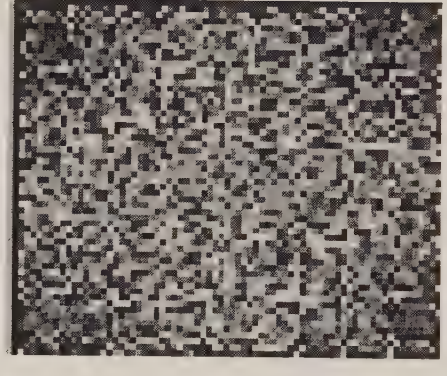
(b)



(f)

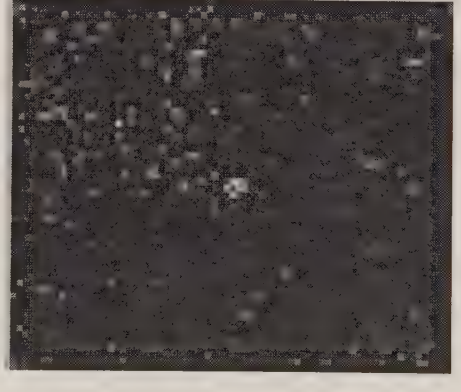


(c)



(g)

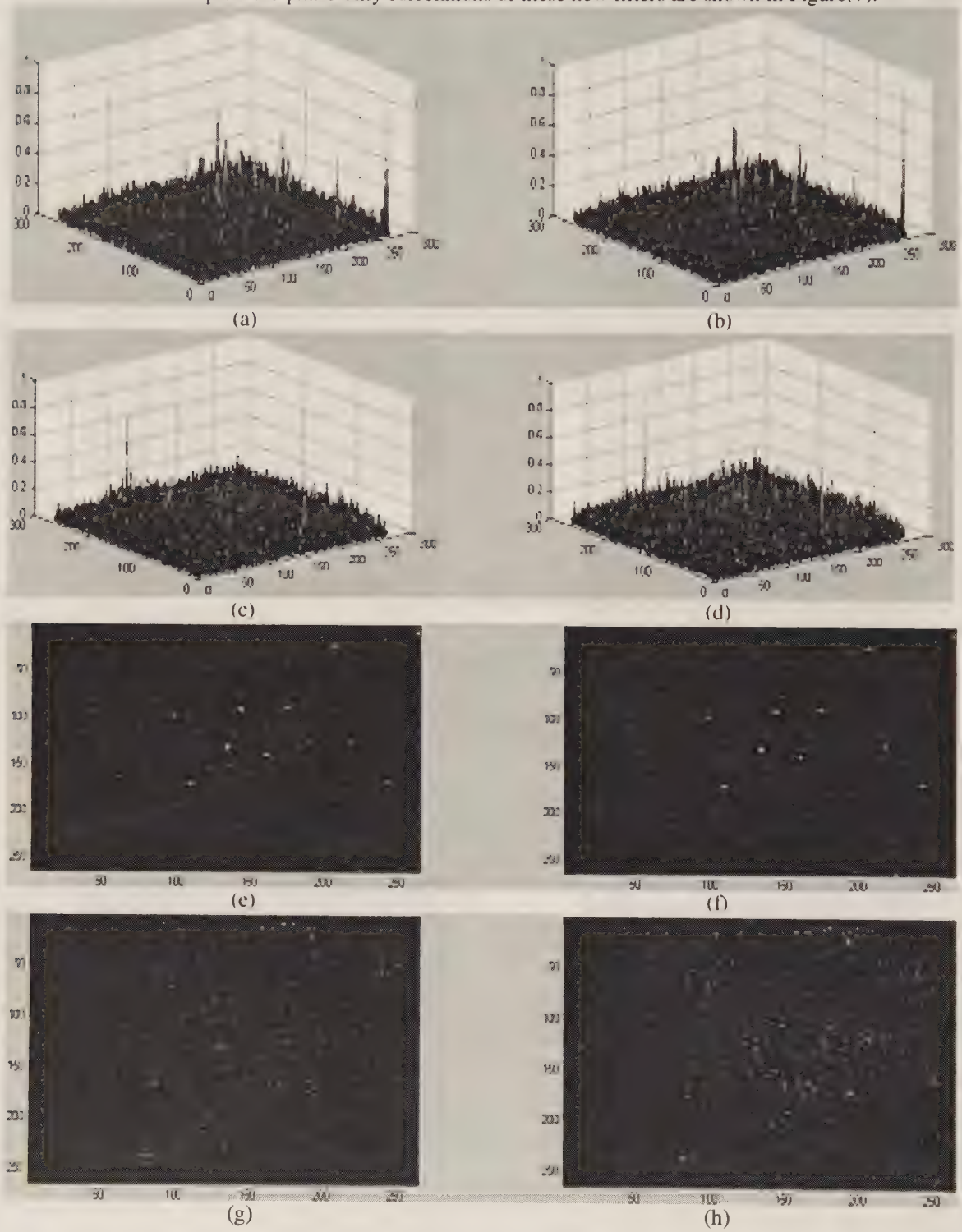
(d)



(h)

Figure (6) The (a)&(c) and (b)&(d) show the magnitudes of the transfer function and impulse response for the KLPSDF, respectively, for each of the two sets of training images, parts (e)&(g) and (f)&(h), show the KLPSDF phase for the phase-only implementation, and the impulse response for the phase only filter implementation for each of the two sets of training images, respectively.

The results of the complex and phase-only correlations of these new filters are shown in Figure(7).



Figure(7) The correlation results are shown in (a)&(b) and (e)&(f) for the complex and phase only-implementation of the each of the top/bottom-views and (c)&(d) and (g)&(h) for the side-views of the protein structures, respectively.

We observed the correlation plane results and were able to identify the protein structures. This was done through further processing of each of the correlation planes and the relative comparison of these planes. In the case of the top/bottom view results, the complex implementation of the PSDF was able to identify all 12 protein structures with no false detection, this was also the case for the POF implementation of the PSDF. The side view implementation of these filters presented more of a challenge. Only, seven of the structures were identified with no false detection for the complex PSDF implementation, and eight structures were identified for the POF implementation. The KLPSDF complex and phase only implementation revealed 10 of the 12 top views of the protein structures in the region under consideration of the test pattern, though, the complex implementation would yield the 12 structures in the top view with some more post processing. We think the reason for this reduction in performance has to due with the redundancy of the information contents in the first set of training images. The KLPSDF will do better for training images that are different in steady rate. In other words, if the training images have the same amount of variation, the more the better for this filter, from one training image to the next, then the KLPSDF will do a better job in finding that class of objects as the KL transform will encode a similar amount of energy in the filter for each of the variations. The KLPSDF implementation for the side view of the proteins had better results than the PSDF implementation. This is to be expected, as the variation between the training images is significantly more in the side view training images, than those in the top/bottom view images. The KLPSDF picked ten of the twenty side views in the phase only implementation. The complex implementation only picked five of the side views of the proteins, this is primarily due t the lack of energy contents in these training images, the information content was there, as we viewed in the POF implementation results.

VII. Conclusions

In this work we combined the KLT and the PSDF filters to implement a new class of filters. The KLPSDF allows for encoding of more information in the filters. This type of information encoding requires consideration of the amount of energy in the training images as the low-frequency content information will have more energy associated with them. The best application of these filters will find themselves in scenarios where there is more variation between different training images, and where there is a large amount of high frequency content in each of the training images. Further work in this area should include a method to spread the information contents of the training images evenly, in terms of energy, across the training images and perhaps across the test images. This may be possible by carefully using a PRN code spread the information contents to other frequencies with in a given band.

VIII. Acknowledgment

The authors would like acknowledge their discussions with Dr. Grant Jensen at the Jensen Laboratory at the California Institute of Technology, in Pasadena CA, on the topic of Automated Particle Picking, and thank him for making himself available to provide fruitful discussions and explanations on this very interesting topic.

References

[1] Riasati, V. R., "Projection-Slice Synthetic Discriminant Functions for Optical Pattern Recognition," Ph. D Dissertation, University of Alabama in Huntsville, ECE Dept. .

[2] Riasati, V. R., Abushagur, M. A. G., "Projection-Slice Synthetic Discriminant Functions for Optical Pattern Recognition", Applied Optics, Vol. 36, No. 14, pp 3022-3034, 10 May 1997.

[3] Hester C. F. and Casasent D., "Multivariant technique for multiclass pattern recognition." Appl. Opt. 19, 1758-1761 (1980).

[4] Caulfield H. J. and Maloney V. T. "Improve discrimination in optical character recognition," App. Opt. 8, 2354-2356 (1969).

[5] Bracewell, R. N., "Strip integration in radio astronomy," Aust. J. Phys., Vol. 9, pp 198-217, (1956).

[6] Dudgeon, D., and Mersereau, R. M., Multidimensional Digital Signal Processing, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, (1984).

[7] Gonzalez, Rafeal C. Digital Image Processing. New Jersey: Addison-Wesley, 2002.

[8] Automated Molecular Imaging: Automatic Particle Picking page at: http://ami.scripps.edu/prtl_data/klh/klh_1k/index.htm,

[9] Mersereau, R. M., "Direct Fourier transform techniques in 3-D image reconstruction, " Computers in Biology and Medicine, Vol. 6, pp 247-58, (1976).

[10] Riasati, V. R., Abushagur, M. A. G., Gregory, D., "Phase-Only Projection-Slice Synthetic Discriminant Functions for Invariant Optical Pattern Recognition", SPIE Conf., April 1997, Orlando, FL.

[11] Feature extraction with wavelet transforms for implementation of the projection-slice filter, Riasati, Vahid R., Gregory, Don A., Proc. SPIE Vol. 3715, p. 259-271, March 1999.

Optical Implementation of Image Encryption Using The Zero-Order Phase-Contrast Technique

Jose C. Pizolato Jr. and Luiz Goncalves Neto*
Depto. Eng. Elet. - EESC, University of São Paulo - USP, Av. Dr. Carlos Botelho 1465,
13560-970 São Carlos - SP - Brazil,

ABSTRACT

An encryption/decryption scheme based on a new Phase-Contrast Technique, without the use of a phase-changing plate (phase dielectric dot) on the Fourier plane of a 4f optical correlator is proposed. The encryption of a gray level image is achieved by multiplying the phase distribution obtained directly from the gray level image by a random phase distribution. The encoding is obtained without any iterative calculation to generate the encrypted phase-only mask. The robustness of the encoding is assured by the non-linearities intrinsic to the phase-contrast method and the random phase distribution used in the encryption process. The advantage of this method is the easy scheme to recover the gray level information from the decrypted phase-only mask applying the proposed Zero-Order Phase-contrast Technique.

Keywords: Optical Image Encryption, Zero-Order Phase-contrast Technique, Phase Masks, Fourier Phase-contrast Hologram, Spatial Light Modulator.

1. INTRODUCTION

Previous works have proposed the use of the phase contrast technique[1] and random phase distributions to encrypt images in phase-only masks[2,3]. In reference 2, the encryption of a gray level image a(x,y) ($0 \le |a(x,y)| \le 1$) is achieved by multiplying a phase distribution $F(x,y) = \exp[ja(x,y)]$ obtained directly from a gray level image a(x,y) by a random phase distribution $\exp[j2\pi b(x,y)]$, resulting in an encrypted phase-only mask $F_e(x,y) = F(x,y) \exp[j2\pi b(x,y)]$, where b(x,y) is a uniform random distribution in the interval [0,1]. The robustness of the encoding is assured by the non-linearities intrinsic to the phase contrast method and the bandwidth of the random phase distribution[2]. The content of the phase mask $F_e(x,y)$ cannot be determined by light intensity detectors and it is also extremely complex to recover the encrypted image by blind deconvolution, since the necessary condition to assure linearity between the phase shift and image intensity is not respected. Because of the wide bandwidth of the random phase distribution multiplied to the phase image F(x,y) and the finite aperture intrinsic to any optical system, severe information loss will also occur and the reconstructed image will include such strong speckle noise so that it becomes undistinguishable[2]. Without knowing the key mask $\exp[-j2\pi b(x,y)]$, it is impossible to decode the phase mask $F_e(x,y)$. The intensity of the encrypted image can be optically recovered by applying the phase contrast technique to the phase mask $F_e(x,y)$ multiplied by the phase mask key $\exp[j2\pi b(x,y)]$.

To assure a robust optical implementation of the decryption process, in this work we propose an encryption/decryption scheme based on the Zero-Order Phase-Contrast Technique. In this new technique, the intensity distribution is recovered by using a spatial filter allowing only the transmission of the zero order in the Fourier plane. The advantage of this method is the easy scheme to recover the gray level information from the decrypted phase-only mask. The robustness of the encoding is also assured by the non-linearities intrinsic to the new phase-contrast method and the random phase distribution used in the encryption process. The encoding of the image is obtained without any iterative calculation to generate the encrypted phase-only mask.

* LGNETO@sel.eesc.usp.br

Optical Pattern Recognition XV, edited by David P. Casasent, Tien-Hsin Chao, Proc. of SPIE Vol. 5437 (SPIE, Bellingham, WA, 2004) · 0277-786X/04/$15 · doi: 10.1117/12.542461

The phase-only mask containing the encrypted image can be used as a device to verify the authenticity of cards or any other documents. The advantages in use phase-only masks lies in the impossibility of duplicate the information of the phase distribution using common image replication methods.

2. THE ZERO-ORDER PHASE-CONTRAST TECHNIQUE

The Zero-Order Phase-Contrast Technique is based on an image-forming system like the 4f optical correlator, shown in Fig. 1. In this new technique, the intensity distribution is recovered from the phase information by using a spatial filter on the Fourier plane, allowing only the transmission of the zero order. A magnification of unity is assumed and the finite extent of the entrance and exit pupil is not considered.

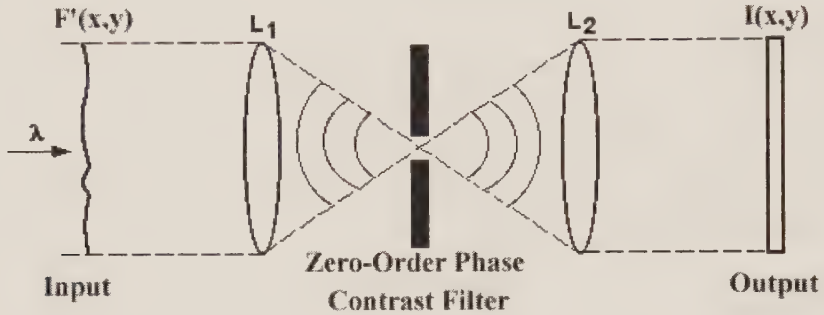


Figure 1: . System that forms the Zero-Order Phase-Contrast Technique, based on an image-forming system like the 4f optical correlator.

Suppose now a phase object F'(x,y), formed by m×n square pixels of size M×N, generated from a gray level image a(x,y) of m×n square pixels, as shown in figure 2. The phase object F'(x,y) is placed in the input plane and is coherently illuminated. Each pixel is composed by two regions that modulates different values of phase. One region is formed considering the information of a(x,y), and the other by the phase value exp[jπ]. The complex transmittance of F'(x,y) is described by the equation:

$$F'(x,y) = \left\{ exp[ja(x,y)]III\left(\frac{x}{N}, \frac{y}{M} - \frac{M}{2}\right) + exp[j\pi]III\left(\frac{x}{N}, \frac{y}{M} + \frac{M}{2}\right)\right\} * rect\left(\frac{x}{N}, \frac{2y}{M}\right) \quad (1)$$

Neglecting the terms of order $a(x,y)^2$ and higher, the amplitude transmittance F'(x,y) can be approximated by the equation

$$F'(x,y) \cong \left\{ [1 + ja(x,y)]III\left(\frac{x}{N}, \frac{y}{M} - \frac{M}{2}\right) - III\left(\frac{x}{N}, \frac{y}{M} + \frac{M}{2}\right)\right\} * rect\left(\frac{x}{N}, \frac{2y}{M}\right) \quad (2)$$

The image produced in the output plane without the spatial filter on the Fourier plane is written by

$$I(x,y) \cong \left| \left\{ [1 + ja(x,y)]III\left(\frac{x}{N}, \frac{y}{M} - \frac{M}{2}\right) - III\left(\frac{x}{N}, \frac{y}{M} + \frac{M}{2}\right)\right\} * rect\left(\frac{x}{N}, \frac{2y}{M}\right)\right|^2 \cong 1 \quad (3)$$

The diffracted light is not observed because it is in phase quadrature with the background. If this phase quadrature term is isolated by using a spatial filter on the Fourier plane, intensity variations will occur and the transparent object could be observed. In our optical simulations, the spatial filter is formed by a metal plate with a square aperture allowing only

the transmission of the zero order in the Fourier plane. In figure 1, the complex distribution of light onto the filter is given by

$$S'(u,v) \cong \left\{ jA(u,v) * \left[III(Nu,Mv)\exp(-j\pi Mv) \right] + III(Nu,Mv)\exp(-j\pi Mv) - III(Nu,Mv)\exp(j\pi Mv) \right\} \times sinc\left(Nu, \frac{M}{2} v \right) \tag{4}$$

The complex distribution of light just behind the filter (assuming a square aperture) is given by

$$S'(u,v) \cong \left\{ jA(u,v) * \left[III(Nu,Mv)\exp(-j\pi Mv) \right] + III(Nu,Mv)\exp(-j\pi Mv) - III(Nu,Mv)\exp(j\pi Mv) \right\} \times sinc\left(Nu, \frac{M}{2} v \right) rect(Nu,Mv) \tag{5}$$

Considering only the zero order terms of Equation 5 we have

$$S'(u,v) \cong \left[jA(u,v) * \delta(u,v) + \delta(u,v) - \delta(u,v) \right] sinc\left(Nu, \frac{M}{2} v \right) = jA(u,v)\, sinc\left(Nu, \frac{M}{2} v \right) \cong jA(u,v) \tag{6}$$

In figure 1, the lens L_2 performs the inverse Fourier transform of A(u,v), resulting in the intensity distribution

$$I(x,y) = a(x,y)^2 \tag{7}$$

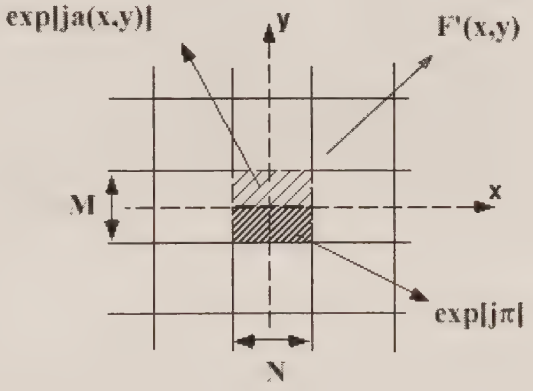
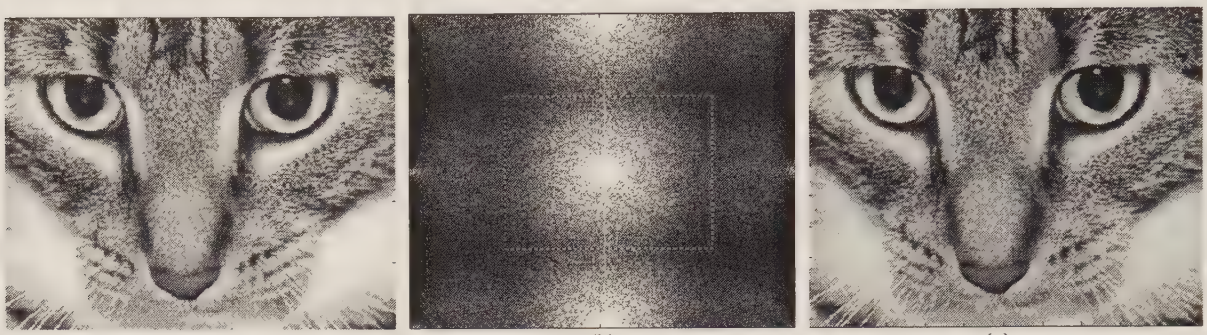


Figure 2: Phase object used to implement the Zero-Order Phase-Contrast Technique. F'(x,y) has m×n square pixels of size M×N, generated from a gray level image a(x,y) of m×n square pixels. Each pixel is composed by two regions that modulates different values of phase. One region is formed considering the information of a(x,y), and other by the phase value exp[jπ].

Fig. 3 shows the computer simulation of the proposed Zero-Order Phase-contrast Technique. The square root of an image with 256×256 pixels, Fig. 3(a), is used to generate the amplitude information a(x,y). A 512×512 matrix with the phase information F'(x,y) is generated from a(x,y) as described in Fig. 2 and Equation 2. A 512×512 matrix with the image of the amplitude information S'(u,v) in Fourier plane is shown in Fig. 3(b). Only the delimited area in Fig. 3(b) is

used to recover the original information a(x,y) using the inverse Fourier transform. A 512×512 matrix with the intensity of the recovered distribution is shown in Fig. 3(c).



(a) (b) (c)

Fig. 3: Computer simulation of the Zero-Order Phase-contrast Technique. (a) figure used to generate the amplitude information a(x,y). The 512×512 matrix with the phase information F'(x,y) is formed from a(x,y) as described in Fig. 2 and Equation 2. (b) 512×512 matrix with the image of the amplitude information |S'(u,v)| in Fourier plane. Only the delimited area is used to recover the original information a(x,y) from the inverse Fourier transform. (c) 512×512 matrix with the intensity of the recovered distribution $I(x,y)=|a(x,y)|^2$.

2. IMAGE ENCRYPTION AND DECRYPTION APPLYING THE ZERO-ORDER PHASE-CONTRAST TECHNIQUE

The encoding method used to encrypt a gray level image is similar to the technique described in reference 2. Firstly the distribution F'(x,y) is generated from the gray level image a(x,y) ($0 \le |a(x,y)| \le 1$) using Equation 1. The encryption of the phase distribution F'(x,y) is achieved by multiplying a random phase distribution exp[j2πb(x,y)], resulting in a encrypted phase-only mask $F'_e(x,y) = F'(x,y)\ \exp[j2\pi b(x,y)]$, where b(x,y) is a uniform random distribution in the interval [0,1]. Without knowing the key mask exp[-j2πb(x,y)], it is impossible to decode the phase mask $F'_e(x,y)$. The intensity of the encrypted image could be optically recovered by applying the Zero-Order Phase Contrast technique to the phase mask $F'_e(x,y)$ multiplied by the phase mask key exp[-j2πb(x,y)]:

$$F'_d = F'_e \cdot exp\left[-j2\pi b(x,y)\right] \tag{8}$$

The experimental set-up in Fig. 4 is used for implementing the decryption applying the Zero-Order Phase-Contrast Technique. A He-Ne laser is used as the coherent light source. An expanded beam of light is incident on two phase Spatial Light Modulator - SLM carefully aligned. The SLM's are two *Seiko-Epson InFocus TVT-6000* operating in phase mostly regime mode between 0 and 1,5π. The SLM's are characterized using the technique described in reference 4. A critical point in the system described in Figure 4 is the alignment between the SLM's. This problem can be minimized by smoothing the information to be modulated. The SLM's are used to simulate the multiplication of the encrypted phase mask to the phase mask key. The first SLM modulates the encrypted information $F'_e(x,y)$, and the second the phase mask key exp[-j2πb(x,y)]. The lens L1 implements the optical Fourier transform of the final product in the input of a 4f optical correlator. In the Fourier plane of the correlator, only the zero order component diffraction is transmitted. The lens L2 implements the inverse Fourier transform of the components filtered. The recovered information imaged onto a CCD camera and is shown in Fig. 5.

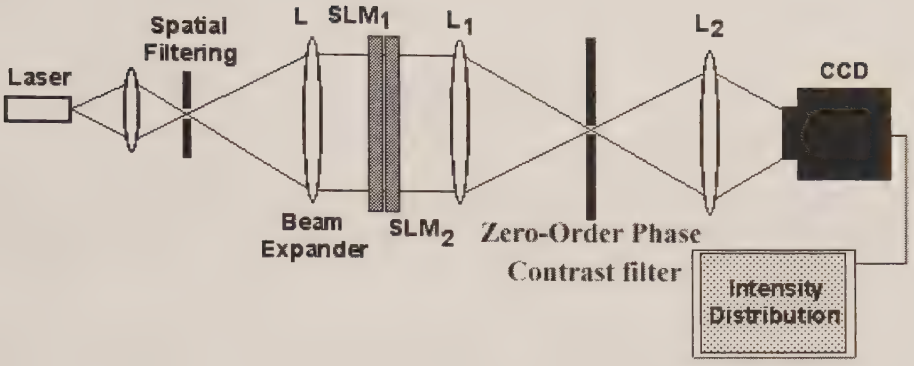


Figure 4: Experimental set-up used for implementing the Zero-Order Phase-Contrast Technique. A He-Ne laser is used as the coherent light source. An expanded beam of light is incident on two phase Spatial Light Modulator - SLM carefully aligned. The SLM's are two *Seiko-Epson InFocus TVT-6000* operating in phase mostly regime mode between 0 and 1,5π. The first SLM modulates the encrypted information $F'_e(x,y)$, and the second the phase mask key $\exp[-j2\pi b(x,y)]$.

Figure 5: The recovered information imaged onto a CCD camera.

3. CONCLUSION

A simple technique to encrypt/decrypt an image using the Zero-Order Phase Contrast Method is proposed. This new method assures a robust image reconstruction when the encrypted phase mask multiplied by the phase key is placed in the input of a 4f optical correlator with a zero-order pass filter on the Fourier plane.

ACKNOWLEDGEMENTS

The authors are supported by FAPESP - Brazil, CNPq - Brazil and FINEP - Brazil.

REFERENCES

1. J. W. Goodman, "Introduction to Fourier optics,,, 48-54, McGraw-Hill, (1968).

2. L. G. Neto, "Implementation of image encryption using the phase-contrast technique,,, in Optical Pattern Recognition IX, Proceedings of SPIE, Vol. 3386, p. 284-290, D. P. Casassent, Tien-Hsin Chao; Eds., Orlando'98 SPIE's AeroSense Symposium, April 12-17, 1998.

3. J. Glueckstad, P. C. Mogensen, R. L. Eriksen, "New phase contrast methods for optical processing", in Optical Pattern Recognition XIII, Proc. SPIE Vol. 4734, p. 13-22, David P. Casasent; Tien-Hsin Chao; Eds., Orlando 2002 SPIE's AeroSense Symposium, April 1-5, 2002.

4. Luiz Goncalves Neto, Danny Roberge and Yunlong Sheng, "Programmable optical phase-mostly holograms with coupled-mode modulation liquid crystal television,,, Applied Optics 34, 11, 1944-1950, (1995).

5. Luiz Goncalves Neto, Danny Roberge and Yunlong Sheng, "Full range continuous complex modulation using two coupled-mode liquid crystal televisions,,, Applied Optics, Vo. 35, No. 23, p. 4567 4576, (1996).

Classification and rejection of MSTAR data

Lakshmi Ramamoorthy and David Casasent
Carnegie Mellon University, Dept. of Electrical and Computer Engineering, Pittsburgh, PA 15213

ABSTRACT

Classification and rejection tests were performed on the 10-class MSTAR database. To compare our performance, we first summarize relevant prior work. Shift-invariant magnitude Fourier transform (FT) features were used in the feature space trajectory (FST) classifier to classify the 10-class MSTAR data with variants and to reject confuser images and clutter chips. No prior work has addressed this. Implication of various SAR preprocessing on performance is addressed. In confuser rejection, 2 standard confusers (used in prior work) and 2 new confusers are addressed. We are the first to extend confuser rejection tests to 8 target classification with variants and 2 confuser rejection. Finally, clutter rejection while classifying the 10 targets is addressed.

Keywords: clutter rejection, confuser rejection, SAR classification and variant recognition.

1. INTRODUCTION

The Synthetic Aperture Radar (SAR) database used for classification was obtained through the Moving and Stationary Target Acquisition (MSTAR) program. The publicly released portion of the MSTAR database contains SAR images of 10 military vehicles (Table 1.1). Each of the targets has views at 15° and 17° depression angles. The 17° data are used for training and the 15° data for testing. There are about 190-300 different azimuth versions of each target at each depression angle. Table 1.1 lists the major testing and training sets available. Other T72 variants and targets at 30° and 45° depression angles are available, but the ones noted are the major ones used by researchers. Among the 10 different targets in Table 1.1, there are 2 Tanks, 4 Armored Personnel Carriers (APCs), a Self-propelled Artillery/ Rocket Launcher and an Air Defense Unit (ADU). The D7 and ZIL131 are a bulldozer and a truck respectively. They are used as one set of confusers (non-military vehicles) to be rejected. No training set images of any confusers are allowed.

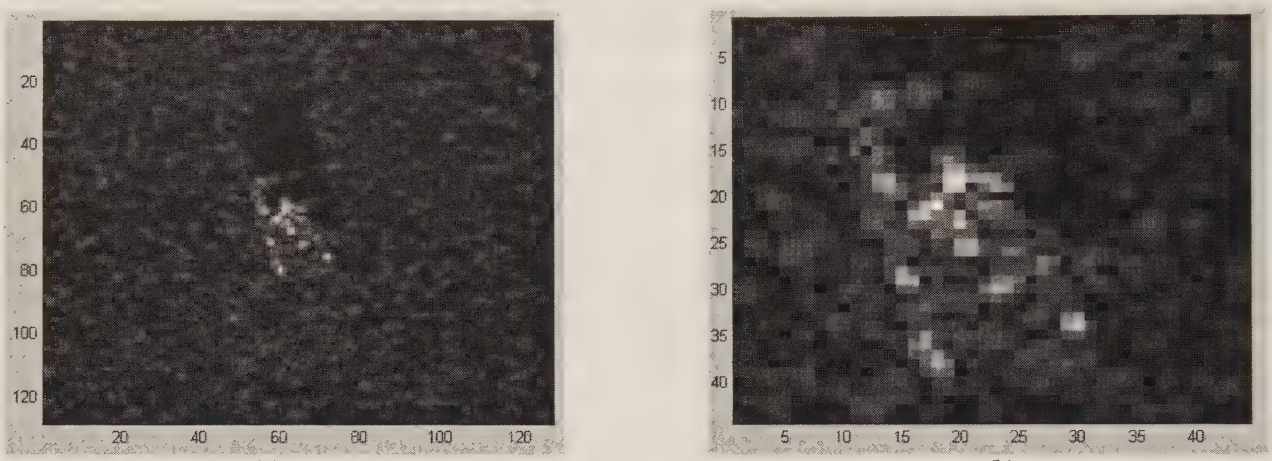
Vehicle Type	Vehicle-variant	Depression Angle (Degrees)	No. of aspect views available in training set	No. of aspect views available in test set
APC	BMP2-9563	15,17	233	195
	9566	15	-	196
	c21	15	-	196
	BTR70	15,17	233	196
	BTR60	15,17	256	195
	BRDM_2	15,17	298	274
Tank	T72-132	15,17	232	196
	812	15	-	195
	s7	15	-	191
	T62	15,17	299	273
Rocket Launcher	2S1	15, 17	299	274
ADU	ZSU_23_4	15,17	299	274
(Our confusers)				
Bulldozer	D7	15,17	299	274
Truck	ZIL131	15, 17	299	274

Table 1.1. List of major MSTAR training and test targets, used by us and other researchers, their depression angles and the number of aspect views available

Optical Pattern Recognition XV, edited by David P. Casasent, Tien-Hsin Chao, Proc. of SPIE Vol. 5437 (SPIE, Bellingham, WA, 2004) · 0277-786X/04/$15 · doi: 10.1117/12.548072

Two of these targets (BMP2 and T72) have version variants as well. The goal of classification tests is to recognize objects with a 2° difference in depression angle and to recognize variants without having trained on them. The clutter database consists of 100 clutter scenes, each of size 1784×1478 pixels, at a depression angle of 15°. The clutter scenes are mainly trees etc. In earlier work [6], 1616 regions of interest were obtained from these clutter images using a screener. We use the central 44×44 pixel region of these same [6] 1616 regions as our clutter to be rejected.

In confuser rejection tests, two of the objects (D7 and ZIL131), which are not high demand military targets, are used as new *confuser targets*. We also use the 2S1 and D7 as another pair of confusers, since other researchers did. The classifier should reject confusers as not being targets of interest (military vehicles) without training on them. In our *non-variant tests*, we train on the 17° depression angle views and we test on the 15° depression angle views. In our *variant tests*, we test on the 15° depression angle aspect views and we include all aspect views of the 3 variant versions of the BMP2 (9563, 9566, C21) and the 3 variants of the T72 (132, 812, S7) in the test set. Only the first variants of the T72 and BMP2 at 17° depression angle are present in the training set. Our training set data are taken from the data in Table 1.1. *No confuser images are included in the training phase.* Table 1.1 also lists the 15° depression angle test set we used in our tests when variants are included.



(a) (b)

Figure 1.1. Target chip. (a) Full size, (b) Central region used in the classification tests

Fig. 1.1a depicts a target chip at a 17° depression and a 346° azimuth angle. The target chips in the database are of sizes ranging from 128×128 pixels to 192×193 pixels. They consist of the target that is nearly in the center, its shadow and background clutter. We cropped all target chips to the central 44×44 pixels (Fig. 1.1b); this retains the target and a portion of its shadow. In prior work, various target chip sizes have been used; typically 64×64 pixels [6], 96×96 pixels [4], etc. The background around the object has a fairly low intensity; it may affect performance of some classifiers, but is not easily removed without modifying the object; it is thus present in training and test data. Thus, smaller target chips were used by us. Although the entire target is present in these chips, it is not always in the center. *This is of vital concern. When the feature space used for classification does not use shift-invariant features, centering of the target in ROIs is a critical issue.* About 6.5% of the 10-class test set (without variants) are noticeably shifted and will be misclassified unless shift-invariant features are used. We did not find an adequate method to center the objects in the database (even by correlation).

Log-scaled and energy normalized images are used with our features. We bias the image (by adding 1) before log scaling, so it contains only positive values. Some of the images have very high pixel values on the targets, we log-scale the images so that these pixels do not dominate the features. For example, one test image without log scaling has a range of pixel values from 0.0018 to 11.1192; when this image is log-scaled, its pixel range varies from 0.0008 to 1.0835. Log-scaling thus reduced high intensity pixels. Without log-scaling, several pixels in each image have very large values and thus the image appears nearly black. With log-scaling, more of the object pixels are apparent and the contrast of the object is better. Log-scaled images are further normalized to unit energy and then the features are extracted for classification. We used 44×44 pixel magnitude Fourier transform (FT) features (since they are shift invariant) of log-scaled and energy normalized chips. The dimensions of the features space can be reduced by wedge ring detector [15] etc sampling. Sect. 3 describes the feature space trajectory classifier used. Sect. 4 presents test results.

A brief note about the performance measures used follows. P_D is the probability of detection of the targets (regardless of whether classification is correct), P_C denotes the probability of classification of the targets over the entire test set (it includes non-detected targets also), P_{FA} is the probability that a confuser is not rejected and P_{CFA} is the probability that a clutter chip is not rejected. There are two types of misclassification, targets that are rejected as clutter or confusers due to the threshold used and targets that are misclassified as some other target. Equal error rate (EER) is an operating point on the receiver operating characteristics (ROC) curve at which $1-P_C$ and P_{FA} are equal. We also give P_{CM}, the maximum P_C, when no rejection is considered and all inputs are classified.

2. PRIOR MSTAR CLASSIFICATION WORK

In the MSTAR program, performance was given at a detection rate P_D=0.9. Scores are given for P_{CC}, the number of detected objects correctly classified (conditional classification rate). The *P_C rate we use and note for prior work is P_C=$P_{CC}P_D$* and we calculate P_C for all objects (not just those that are detected, P_{CC} is this). Many prior classifiers considered did not perform well, especially in rejection P_{FA}, since they were not designed for rejection. As the threshold T is varied to detect more objects, clearly more non-objects will be detected and P_{FA} becomes worse, when similar algorithms are used for detection and classification. If the normal ROC performance is not good, then varying T to increase P_D to 0.9 will include many non-objects. Over 10% of the test set objects look different from those in the training set; thus, the P_D=0.9 operating point may not be a good one.

All results in this section used MSTAR data *with images at the 17° depression angle for the training set and images at the 15° depression angle for testing.* We first discuss prior results using the MSTAR 3-class database. This consists of the BMP2, BTR70 and T72 vehicles. When tested with variants, the system must recognize the two variants of the BMP2 and the two variants of the T72 at a 15° depression angle in the test set. The confusers to be rejected were the 2S1 (a rocket launcher) and the D7 (a bulldozer) at a 15° depression angle, even though one is a military vehicle. No variant or confuser images are present in the training set. In our work, we used the same confusers and we also considered rejecting the D7 and the ZIL131 (a truck), as they are not military vehicles.

Some prior work noted different methods to estimate the pose of the target [2, 10] and assumed pose was known prior to classification. This significantly improves performance for many classifiers, since only data in the correct aspect bin is considered. However, we feel that such methods are not realistic, *since pose estimation of SAR objects is known [13] to be poor with ±20° estimation errors.* Distortion invariant filter (DIF) work [2, 3] also noted that the pose of the test input can be estimated. It seems that in 3-class tests [2, 3], 16 MACH filters for 16 different angular aspect bins were used in the first stage, and the one with the peak to sidelobe ratio (PSR) above a threshold was used as a pose estimate that was used to restrict the second stage aspect bin DCCF classifier filters used. On 10-class data, in [6], two stages of DIFs were not used. If test pose is assumed known, performance will noticeably improve, as fewer final stage filters are considered. No data is available on how well the pose estimates provided by MACH filters performed.

Prior results on the 3-class MSTAR database are now summarized. In one template matching scheme [14], a 96×64 pixel section of the target chip was used. A separate template was formed for every angular aspect bin (typically 5° bins are used). Each template was simply the average of all training images in the associated aspect range. This is about 72 templates per object or 216 templates for the 3-class database. This seems excessive and requires too many calculations to be of use. The P_C score was 92.67% with variants. The P_C scores drop to approximately 91%, 90%, 85% and 75% if fewer templates are used, being formed for bins of 10°, 15°, 30° and 45° angular increments respectively. They did not consider confuser rejection tests; this would greatly reduce P_C. Shifting of the test inputs for classification was also not done. The effect of signal-to-clutter ratio (SCR) of the target chips on performance was also addressed [14]. P_{CC} decreased gradually from 92.67% to about 84% as the SCR decreased from 8.6dB (the nominal value in MSTAR) to -10 dB (using synthetic Rayleigh clutter). Beyond -10dB, the decrease in P_{CC} was almost linear; it reached 40% for an SCR of -20 dB.

This work [14] is interesting, since the effect of several different transforms on the gray-scale pixel levels in the images were considered: linear, logarithm and power (raising each pixel to different powers α=0.5, 2, etc) transforms. Each transformation was applied to the training and test set. All templates and all transformed test set chips were also normalized to have zero mean and unit variance. We feel that the use of zero mean chips and templates can cause problems, since the background pixels around the targets now contribute significantly to the matching scores (there are many background pixels with the large templates used and the background values are non-zero). The use of unit variance is similar to our energy normalization; however, we do not produce zero mean images and we use a much smaller target chip (44×44 versus 96×96 pixels). The log transform tends to improve the overall target image. If α=2 in the power

transform, this can enhance bright scatterers. If α=0.5, it can improve the overall contrast and shape information and yields results similar to that of the log transform. We feel that powers α greater than 1 should result in poorer performance, since the same pixels are not necessarily high for a depression angle change of 2° or for a small aspect view change. At the nominal SCR, they [14] found that the log transform and the power transform with α=0.5 performed slightly better (P_C was about 0.5% better) than for the linear transform. Our preprocessing gave much better improvements. The power transform with α=2 performed much worse, as expected; it was about 10% worse, than without any transformation.

Other prior 3-class MSTAR results follow. Distance-based mean squared error (MSE) template matching was also considered [1], where 36 templates were formed per object (each was simply the average of all training images over a range of 10° aspect angle variations). They considered three object classes with their variants and two other confusers (non-object class) to be rejected in the test set. They used two tanks not present in the public MSTAR database as the two confusers to be rejected. There was no discussion of log-scaling or normalization. They achieved an average equal error rate (EER) of 0.3 (and they did not assume that the correct aspect angle was known). This corresponds to P_C=70% at P_{FA}=30%. Our results are much better.

A polynomial support vector machine (SVM) classifier [5] was tested on the 3-class MSTAR database and gave a P_C of 93%. Variants and depression angle differences were considered, however non-object confuser rejection test results were not done. A gaussian SVM [10] was also tested on the 3-class database (with variants) with two confusers. An aspect angle estimation algorithm was noted and claimed to yield results with average estimation errors less than 8 degrees. No details were provided on the pose estimation algorithm. They, thus assumed that the aspect angle of the target is known (this assumption is questionable [13], as are results), and this reference now considered confuser rejection. They obtained P_C=84.9% at P_D=90%, with a large P_{FA}=34.3% of missed non-object rejections. In [5], several other classifiers were also tested and their performance was compared at P_D=90% on the 3-class database with variants and depression angle differences and with 2 confusers, as before the orientation of the objects are assumed known. This makes results suspect. The P_C ranges from 80.4 to 84.6% and the P_{FA} ranges from 46.5 to 72.8% (all P_{FA} values are huge). Thus, we do not detail the classification scores, as all give poor rejection performance. Our P_{FA} results are 18.43% for a P_C of 84.9%, which is about 16% better than the P_{FA} of 34.3% that [5] obtained for a similar P_C of 84.9% for a gaussian SVM.

A continuous prototype method was reported [11]. This is of interest, as it produces a smoother manifold than in the feature space trajectory (FST) classifier we use. It was tested on the three class MSTAR data, variants, depression angle differences and rejection of 2 confusers were addressed. Scores of P_C=87.7% (at P_D=90%) and rejection of two vehicles gave P_{FA}=20.8%. Our performance is comparable; P_C=86.9% at the same P_{FA}=20.8%. But, it is to be noted that this prior work [11] approach assumed that the aspect angle of each test input was known. We find this assumption unfair [13]. Restricting the angular range of the manifolds that had to be considered (by assuming that the aspect is known) should improve performance. The DIFs have a much better P_C, but they did not do confuser rejection.

We now address prior work on the 10-class MSTAR database. There is little such prior work. Another template matching method, referred to as a modified learning vector quantization (LVQ) algorithm [4] was tested on the 10-class MSTAR data. It was tested with 6, 9 and 18 templates per target (each covering a different angular aspect range), which equates to 60, 90 and 180 templates for the full set of 10 targets. The 10-class database was used with nine object classes and one confuser, with nine templates per object. They obtained a good P_C=89% and good rejection P_{FA}=19.3%, but the single confuser to be rejected was used in the training set and in the template synthesis (this is not fair); and it was the class that performed poor in recognition; and no variant recognition was considered.

We now address distortion-invariant filter (DIF) approaches. These results are among the best. For all DIF approaches discussed here, all images were log transformed (after adding a bias of 1) and normalized to unit energy. We assume that shifts are applied in all DIFs. The details of the pose-bin estimation performed remain unclear. Maximum average correlation height (MACH) filters and distance classifier correlation filters (DCCF) [2] for the 3-class SAR problem used 16 aspect bins for each target (16 filters per target, each covering a given aspect view range), a total of 48 filters. If the MACH peak sidelobe ratio (PSR) is above an acceptable threshold, the image is further processed (for classification) by 3-class DCCF filters corresponding to the aspect bins of each of the targets that give a PSR above the threshold for the MACH filters. The correct recognition P_C rate was 97%. Training was done with one variant of each target, and tests were against three variants of the BMP2 and the T72 plus depression angle differences. These scores seem to use only the DCCF filters corresponding to the aspect bin MACH filters with the peak to sidelobe ratio (PSR) above a threshold. Unfortunately, no rejection was considered.

MACH and DCCF correlation filters were also applied to the 10-class MSTAR public database [3]. The 10-classes are all of those in Table 1.1. These correlation filters, again use 16 aspect bins (16 filters) for each target, each

filter covering a given aspect view range of the target. The number of MACH filters used was thus 160. As mentioned above, when the PSR for a MACH filter exceeds an acceptable threshold, the image is further processed with the corresponding DCCF filters (only the DCCF filters corresponding to the aspect bin MACH filters with the peak to sidelobe ratio (PSR) above a threshold were used). Use of one (dominant) eigenvector per filter for the transform that maximizes the performance criterion of the DCCF gave a P_C=96%. The number of DCCF filters needed is 16 aspects × 11 filters = 176 and the total number of filters is 160+176=336. For a higher P_C=98% the total number of filters used was 1744 (with all 9 eigenvectors (evs) per filter used for the optimal transform: the DCCF filters require 16 aspects × 11 filters × 9 evs and the MACH filters still require 160 filters). This is a large number of filters. These P_C scores were very high, but variants were not used (only depression angle differences were present) and confuser or clutter rejection was not considered (since these filters have poor clutter rejection [6]).Thus, these results are not of major use. We obtained a P_{CM} of 97.8% (Sect. 4.2) for the 10-class classification with just depression angle differences and no variants. Our magnitude FT feature performance is thus comparable to that of DIFs.

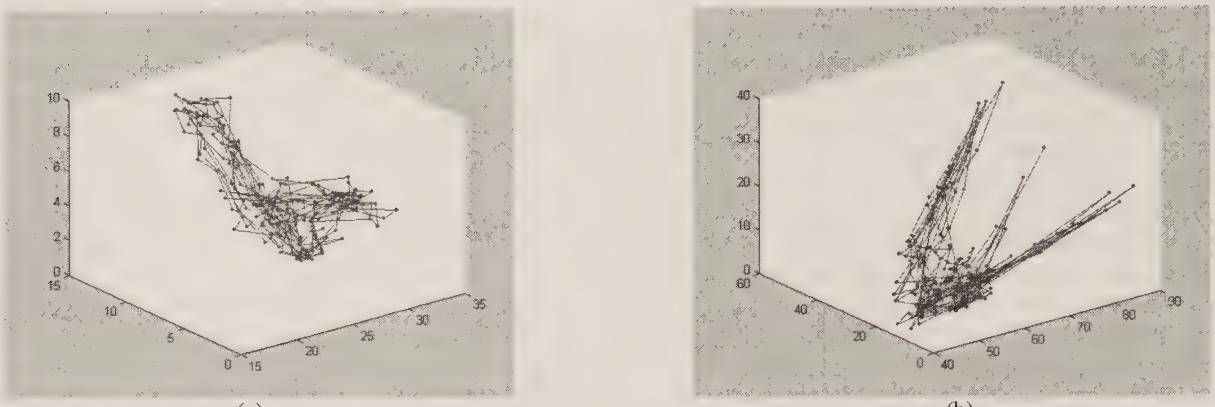
The Extended Maximum Average Correlation Height (EMACH) filter and the Polynomial Distance Classifier Correlation Filter (PDCCF) [6] were developed and then tested on the 10-class MSTAR database. EMACH filters improve the MACH filters by introducing a parameter controlling the relative significance of the average training image in the filter design to achieve better clutter rejection scores. The training set for each object is again divided into sixteen different aspect bins (16 filters were thus used per object). This is the only MSTAR paper that considered clutter rejection; for the 10-class case, 160 EMACH filters were used for detection and rejection of 1616 clutter chips were considered. Confuser false target rejection was not addressed. For depression angle differences only, they achieved P_{CFA}=2.5% (clutter FAs) at P_D=95% and P_{CFA}=1.25% at P_D=90%. With variants, they obtained P_{CFA}=4% at P_D=95% and P_{CFA}=2.5% at P_D=90%. Our P_{CFA} results are much better than these (Table 4.4), and we consider confuser rejection.

In [6], classification using only DCCF, only PDCCF and only EMACH filters is addressed. Nonlinear PDCCF filters yielded improved classification performance. 250 PDCCF filters were used for the 10-class case and achieved an excellent P_C=97.5% versus 96.1% for 160 DCCF filters (for depression angle differences only). Since variants were not always considered in the above data and since confuser rejection was not attempted in any data, comparing results is difficult. When only two variants of the BMP2 and the T72 variants (2 for each of them) were considered, the classification rate using the PDCCF was 83.5% versus 81.2% for DCCF filters. This is P_C for only the variant test inputs. For all 10 classes with variants, we estimated their P_C=93.2%. They did not do pose estimation, but no confuser rejection was considered. Using magnitude FT features, we obtained a comparable P_C of 93.5% for the 10-class case with variants and no confuser rejection (comparable to prior work P_{CM}=93.2% results). We also considered confuser rejection (prior work did not). Of course, P_C performance is reduced when rejection is considered. *Our work is the first to consider the 10-class MSTAR database with variants and confuser rejection.*

3. FEATURE SPACE TRAJECTORY (FST) CLASSIFIER

In the standard nearest neighbor classifier (NNB), each training vector is stored and the test vector is classified as the class of the training vector that is closest in a Euclidean sense. In the FST, we describe each aspect view of an object as a point in a multi-dimensional feature space and we connect the points associated with adjacent aspect views by straight lines to produce a feature space trajectory (FST). Fig. 3.1a shows the FST of the BTR70 SAR object, in a 3-D feature space; we show the first three (or the 3 lowest frequencies) magnitude FT features of the training data of the BTR70 object.

To classify the data using FSTs, an FST is generated for each object, a different feature space is used for each FST. For a test input object aspect view, we calculate its features (in each space) and plot it as a point in each of the FST feature spaces. To classify such a test input, we calculate the Euclidean distances from the test vector to the piecewise linear FST for each object. For detection, we accept the object to be a target if the minimum distance of the test input to any of the FSTs is below a fixed distance threshold. For classification, the test vector is assigned to the class of the closest FST. For rejection, the test input is assigned to the closest FST if its distance is less than some threshold; if it is further than the distance threshold from all FSTs, it is rejected.



(a) (b)

Figure 3.1. FSTs for the BTR70 target using the three dominant magnitude FT feature vectors and different normalizations: (a) Using energy normalized training data for the BTR70 target, (b) without energy normalized training data for the BTR70 target. All cases use log-scaled data.

We now discuss the reasons for use of log-scaled data and energy normalized (enorm) training and test images. The energy per pixel of the different aspect views of the same object varies widely with aspect and for different targets. As a result, each FST is dominated by the large pixel intensities and the shape information is not emphasized. When all chips (training and test set) are normalized to unit energy before being used, this helps intraclass classification, i.e., the various aspect view images of the true class will be more similar. However, we might expect interclass classification and confuser/ clutter rejection to be worse. However, we observed improved rejection of confusers and clutter (Sect. 4.2) using enorm data. This can be attributed to the nature of SAR data. Within a single class of targets, each aspect view image is so different that they lie far apart (in a Euclidean sense) from each other. Fig. 3.1b shows the FST of Fig 3.1a without enorm (but with log-scaling). Note that the axes of the two figures are very different. As a result, the FST without enorm (Fig. 3.1b) occupies a much larger region of the space than it does after enorm (Fig. 3.1a). Many aspect views are much further away from each other when their energies are not normalized (Fig. 3.1b). To obtain good P_C, a larger distance threshold is thus necessary in Fig. 3.1b and this results in worse rejection of confusers and clutter.

4. CLASSIFICATION AND REJECTION WITH MAGNITUDE FT FEATURES

4.1 Introduction

Several different sets of classification tests were performed using different numbers of targets and using different training and test sets, as we now discuss. The details of the 10 MSTAR targets are listed in Table 1.1. Table 1.1 details the number of aspect views of each of the target that is available for training and testing. As seen in the table, the BMP2 and T72 targets have 3 variants each. All results in this section were obtained using magnitude FT features in the FST classifier. No variant or confuser views or clutter chips are used in training.

We now address the contents of the following subsections. In Sect. 4.2, classification performance (without rejection) of the magnitude FT features are considered using different normalizations (with and without energy normalizations); all images are log-scaled and biased. From this we decide on whether or not to use energy normalized images. We present confuser rejection results in Sect. 4.3 and clutter rejection results in Sect. 4.4. We considered 3 cases of confuser rejection testing; the first one (the 3-class database) classifies three targets and their variants (BMP2-9563, BMP2-9566, BMP2-c21, BTR70, T72-132, T72-812 and T72-s7) and rejects two standard confusers (2S1 and D7); the second case classifies the same three targets and their variants and rejects two new confusers (D7 and ZIL131); and the third case classifies eight targets (all of them in Table 1.1, excluding the D7 and ZIL131, which are the two new confusers that we reject). For clutter rejection testing; the ten targets and the variants of the BMP2 and the T72 are classified and the 1616 clutter chips described in Sect. 1 are to be rejected.

4.2 Classification tests

In this section, the classification performance of magnitude FT features with and without energy normalization of the log-scaled and biased MSTAR training and test images is compared. All data are log-scaled. 10-class classification

with and without variants is addressed in this section; i.e., we only consider classification (no rejection distance threshold is used, all inputs are classified). The training set consists of 17° depression angle views of each of the ten objects in Table 1.1 with only the first variant used for objects with variants (BMP2 and T72). The test set consists of all 15° depression angle views of all ten objects and all of their variants noted in Table 1.1. This test is referred to as *10-class classification with variants and without rejection.* We also present results for *10-class classification without variants and without rejection.* We use all aspect views available (Table 1.1) in the training and test set for each of the classes. Table 4.1 shows our test results. We observed that log-scaling the images is essential for classification. P_C obtained without log-scaling was very poor (P_C=9.5%). Hence, we used log-scaled data in all subsequent classification and rejection tests. From Table 4.1, we observe that *the best performance is obtained with energy normalized training and test images.* P_C increased by ~5.5%, for both with and without variants classification, when enorm was introduced. We also observed that log-scaling followed by enorm gave more noticeably better results than just log-scaling for the confuser and clutter rejection tests discussed in Sects. 4.3 and 4.4.

Norms	Total no. of targets	Total no. of misclassified targets	P_C (%)
No enorm and with variants	3203	410	87.2
With enorm and variants	3203	208	93.5
No enorm and no variants	2425	187	92.3
With enorm and no variants	2425	53	97.8

Table 4.1. 10-class classification results using magnitude FT features (no confusers) and FST classifier with and without energy normalization (log-scaled data used for all cases)

We then briefly consider tests with rejection of confusers and clutter. For the confuser rejection tests, classifying three targets and their variants and rejecting two confusers, P_C increased by 21% (from 65.8% to 86.9%) at a P_{FA} of 20.8%, when enorm data was included versus just log-scaled data. Using only log-scaled data, P_C was poor. In clutter rejection tests, classifying ten targets and their variants and rejecting clutter chips, the P_{CFA} decreased by 85.7% (from 85.8% to 0.062%) at a P_D of 90%, when enorm was included versus just log-scaled data. We see that especially for clutter rejection, it is essential to enorm the log-scaled images. Hence, *in all the following subsections of Sect 4, log-scaled and energy normalized image are used.*

Using magnitude FT features and energy normalized training and test images, we obtained P_C=93.5% for the 10-class case with variants and P_C=97.8% for the case without variants. Only DIFs have been tested for classification using the full 10-class MSTAR database. They obtained a comparable P_C of 98% [3] for the 10-class case without variants. Another DIF work [6] obtained a score of 93.2% for the 10-class case with variants (their individual scores for the 10-class case with variants and scores on the variants alone were combined to obtain this score). This is comparable to our score of 93.5% for the case with variants. We note that prior DIF work did not address target rejection as we do in Sect. 4.3.

From Table 4.1, we see that there are only 2.2% errors due to only depression angle differences and more errors (6.5%) due to variant differences. We address this further later.

4.3 Classification and confuser rejection tests

These tests are intended to provide classification with confuser rejection results on several versions of the MSTAR database. Table 4.2 lists the 3 cases we consider and Table 4.3 presents the test set results obtained. The 3-class cases in Table 4.2 consider recognition of three objects: the BMP2 and BTR70 (APCs) and the T72 (Tank). The training set consists of the 17° depression angle views of the first model noted for each of the three objects. No variant or confuser images are included in the training set. The test set consists of all 15° depression angle views of the BTR70 and all 15° depression angle views of all 3 versions (variants) of the other 2 objects. The test set also consists of all 15° depression angle aspect views of the 2 confuser objects noted. All available aspect views of the 3 objects (including variants) and all 548 aspect views of the 2 confusers at the 15° depression angle in Table 1.1 are thus used in the first 3-class confuser tests.

The first 3-class test considered in Table 4.2 is the standard one used by a number of other researchers and thus it was included to allow comparison of our results. Fig. 4.1 shows P_C (percent of test targets correctly classified) versus P_{FA} (percent of confusers not rejected), this ROC (receiver operating characteristic) curve gives test results for the variant case as the threshold T is varied (test inputs > T from all FSTs are rejected as non-objects). The range of T in the ROC

curve in Fig. 4.1 is $8.0 \leq T \leq 16.5$. As we increase T, we recognize more targets, but fewer confusers are rejected (P_{FA} is worse). We note that at the EER=0.075 point we use, T=12.65, and performance is P_C=92.5% and P_{FA}=7.5%.

	Targets	Confusers
3-class standard confuser tests with variants	BMP2 (9563, 9566, c21), BTR70, T72 (132, 812, s7)	2S1,D7
3-class new confuser tests with variants	BMP2 (9563, 9566, c21), BTR70, T72 (132, 812, s7)	D7, ZIL131
8-class confuser tests with variants	BMP2 (9563, 9566, c21), BTR70, T72 (132, 812, s7), BTR60, 2S1, BRDM-2, T62, ZSU-23-4	D7, ZIL131

Table 4.2. Confusers and targets used in the 3-class and 8-class tests with magnitude FT features and FST classifier.

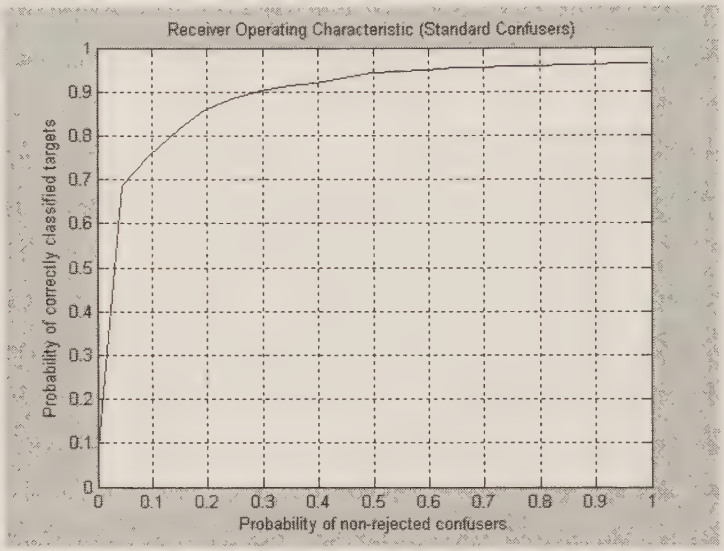


Figure 4.1. ROC classification/confuser rejection curve of P_C for the 3-class case (with variants) vs P_{FA} for the 2 standard confusers, using the FST with magnitude FT features

Table 4.3 lists the score at another point on the ROC (receiver operating characteristic) curve for this first 3 class, 2 confuser case, with variants; the point was chosen, such that we can compare our results with those of prior work. The following performance measures are listed in Table 4.3, P_{CM} (the maximum P_C, when no rejection threshold is used, i.e., when we do not try to reject confusers), P_D, P_C and P_{FA} at a given threshold T (T=13.27 for the first 3-class data). P_D is the percentage of test inputs within T of any FST (even if they are closest to a wrong FST). The 1-P_C target errors are of two types. The total number of target errors is the number rejected as non-objects (those with distance from all FSTs > T) plus the number misclassified (closest to a wrong class FST). Both are listed in Table 4.3.

We first consider the first 3-class case when no target variants are present (only differences in depression angle plus differences in aspect angle are considered). Here P_{CM}=99.8% is very high. At T=13.27, 9 of the 587 targets are missed or misclassified (P_C=98.5%) and only 114 of 548 confusers (P_{FA}=20.8%) are not rejected. Rejecting non-seen objects is difficult. Except for one test input, all the other target errors (8 of 9) without variants are targets too far from FSTs (they are rejected, not misclassified). When variants are included, the maximum P_C obtained drops to P_{CM}=96.3%, which is still good. At the same T=13.27, P_C reduces to 86.9% and we reject all but the same P_{FA}=20.8% of the confuser test inputs (114 of 548); this is the same as before, since T is the same. Of the 1365 target and variant test inputs, 587 had only depression angle differences (no variants) and 778 were variant differences also. We obtained a total of 179 errors (149+30) on all 1365 test images. For only depression angle differences, we found only 9 of 587 or ~ 1.5% errors (from the first test case); however, for variant test inputs, we had 179-9=170 errors out of the 778 variant test inputs or a much larger ~22% error rate (as seen and expected, *most of the variant input errors were variants rejected as non-objects, since they lie further from the true class FST*). Clearly, *recognizing unseen variants of an object is difficult.* Without rejection, we achieved a respectable score of P_{CM}=96.3% correct classification for the variant case.

	3-class (Standard confusers)		3-class (New confusers)		8-class (New confusers)	
	With variants	Without variants	With variants	Without variants	With variants	Without variants
Total number of Targets	1365	587	1365	587	2655	1877
Confusers	548	548	548	548	548	548
Threshold	13.27		13.27		12.72	
No. of targets rejected as non-objects	149	8	149	8	388	125
No. of targets mis-classified	30	1	30	1	103	31
No. of confusers misclassified as target	114		105		114	
P_D (%)	89.1	98.6	89.1	98.6	85.4	93.3
P_C (%)	86.9	98.5	86.9	98.5	81.5	91.7
P_{FA} (%)	20.8	20.8	19.16	19.16	20.8	20.8
P_{CM} (%)	96.3	99.8	96.3	99.8	93.3	97.8

Table 4.3. 3-class and 8 class confuser rejection and target classification test set results using magnitude FT features and FST classifier

As noted, we considered this first 3-class test set to allow comparison to prior work (Sect. 2). On this database, others [10] obtained P_C=80.4-84.9% and a large P_{FA}=34.3-72.8%. Our scores are much better. Distortion-invariant filter work [2, 3] did not attempt rejection (the DIF filters used have poor rejection). The best 3-class prior work scores with variants and rejection [11] (P_C = 87.7% and P_{FA} = 20.8%, hence our T choices, which gave the same P_{FA}=20.8% was used) are comparable to ours (at T=13.27; we have scores of P_C = 86.9% and P_{FA} = 20.8%), but this prior work assumed that the input aspect view was known. As we discussed in Sect. 2, we ignore scores that assumed that the aspect view of the test image was known.

The second 3-class test set (new confusers) noted as "new" in Table 4.2 used 2 different confusers, that are not high-valued military vehicles. This seems a more realistic confuser rejection case. The training sets were the same as above; as were the test sets, except for the 2 confusers used. The same T was used to allow comparison of our results to the first case; hence all results in Table 4.3 (3-class, new) are the same except for P_{FA}. As seen in Table 4.3, in the new confuser set, we fail to reject 105 confusers out of 548. In the standard confuser case (first 3-class test), we fail to reject 114 confusers, or 9 more than in the new case (second 3-class test). Thus, performance is better with these new confusers; this is expected, since they are more realistic, more different objects from the targets. For the scores with variants, we note that at the EER=0.07 versus 0.075 point in the previous case, we use T=12.7, and performance is better with P_C=93% and P_{FA}=7%.

The third test with confusers classifies 8 targets and rejects two confusers. The targets and the confusers used are given in Table 4.2. The 8-class case considers recognition of the BMP2, BTR70, BTR60 and BRDM2 (APCs), the T72 and T62 (Tank), the 2S1 (rocket launcher) and the ZSU23-4 (ADU). The other two classes (D7 and ZIL131) are considered as the confusers to be rejected. The training set consists of all aspect views of the 8 objects at a 17° depression angle, using only the first model noted for the objects with variants (BMP2 and T72) in Table 4.2. The test set consists of all 15° depression angle views of all of the objects mentioned above including all 3 versions (variants) of the BMP2 and the T72. The test set also consists of all 15° depression angle aspect views of the 2 confuser objects noted, i.e., all 548 aspect views of the 2 confusers noted in Table 1.1.

Figure 4.2 shows P_C versus P_{FA} ROC data with variants as the distance threshold is varied ($8.0 \leq T \leq 16$) for this third confuser rejection case. We note that at the EER=0.193 point, we use T=12.68, and performance is P_C=80.7% and P_{FA}=19.3%. Table 4.3 lists the P_{CM} with no rejection, P_D, P_C and P_{FA} scores at another point on the ROC (receiver operating characteristic) curve (T=12.72); the point was chosen, such that we can compare the 8-class case results with those of the first 3-class case (the same P_{FA} for both cases). The scores for the case of no target variants (considering only depression angle variations plus differences in aspect angle) are P_C= 91.7%, P_{FA}=20.8% and an excellent P_{CM}=97.8%. When variants are included, the P_{CM} obtained drops to 93.3%, which is still very good. As seen in the 3-class case also, recognizing variants while rejecting non-seen objects is difficult; P_{CM} reduces from 97.8% to 93.3%. For P_C and confuser rejection, the P_C loss is only ~5.4% from P_C=86.9% (3 classes) to 81.5% (8 classes) at the same rejection

rate of P_{FA}=20.8%, since we did not change T. For only depression angle and aspect angle differences (i.e., without variants), we had 156 of 1877 or ~ 8.3% errors; for only the 778 variant test inputs (2655-1877), 335 of the 778 variant test inputs or ~43% were errors. Thus, variant recognition with rejection is poor. Without rejection, we achieved a P_{CM} of 93.3% for the variant case or P_{CM}=90.75% for just the 778 variant test inputs (with no rejection). We note that most target errors (388 out of 491=388+103) are targets rejected rather than misclassified (with other features, many T72 variants were misclassified as T62 objects, rather than being rejected as occurred here).

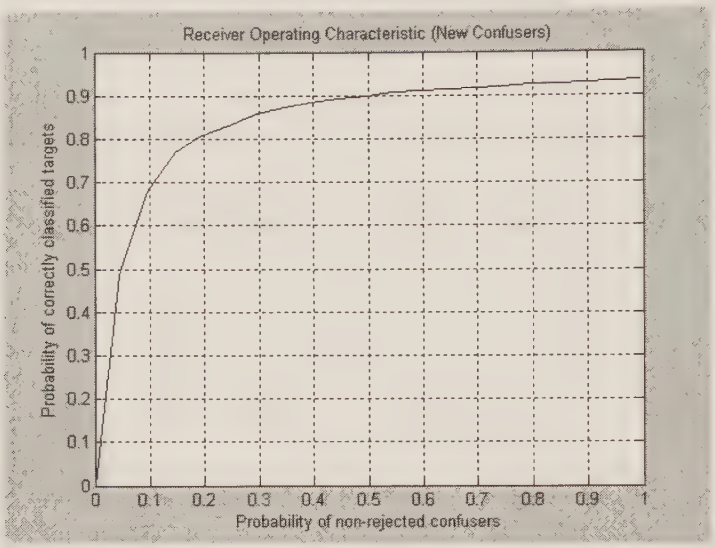


Figure 4.2. ROC classification/confuser rejection curve of P_C for the 8-class case (with variants) vs P_{FA} for the 2 new confusers, using the FST with magnitude FT features

Thus, *we have been the first to extend the 3-class case to the case of 8 classes and 2 confusers*. For the 8-class case with variants, P_C decreased by about 5.4`% and P_{FA} increased by about 1.6%, compared to the 3-class case (both with new variants).

4.4 Clutter Rejection tests

Clutter rejection tests using ten class classification and 1616 clutter chips for rejection was done. The training set consists of the 17° depression angle views of each of the ten objects in Table 1.1 with only the first version used for the 2 objects with variants (BMP2 and T72). The test set consists of all 15° depression angle views of all 10 objects and all of their variants noted in Table 1.1. The clutter data used in testing was described in Sect. 1; there are a total of 1616 clutter chips, each of size 44×44 pixels, at a 15° depression angle.

Figure 4.3 shows the P_C (with variants) versus P_{CFA} (percent of clutter not rejected) ROC data as the threshold is varied ($8 \leq T \leq 23$). We note that at the EER=0.068 point, we use T=14.64, and performance is P_C=93.2% and P_{FA}=6.8%. Figure 4.4 shows the P_D (with no variants) versus P_{CFA} (percent of clutter not rejected) ROC data as the threshold is varied ($8 \leq T \leq 23$). P_D is the percentage of test inputs within T of any FST (even if they are closest to the wrong FST).

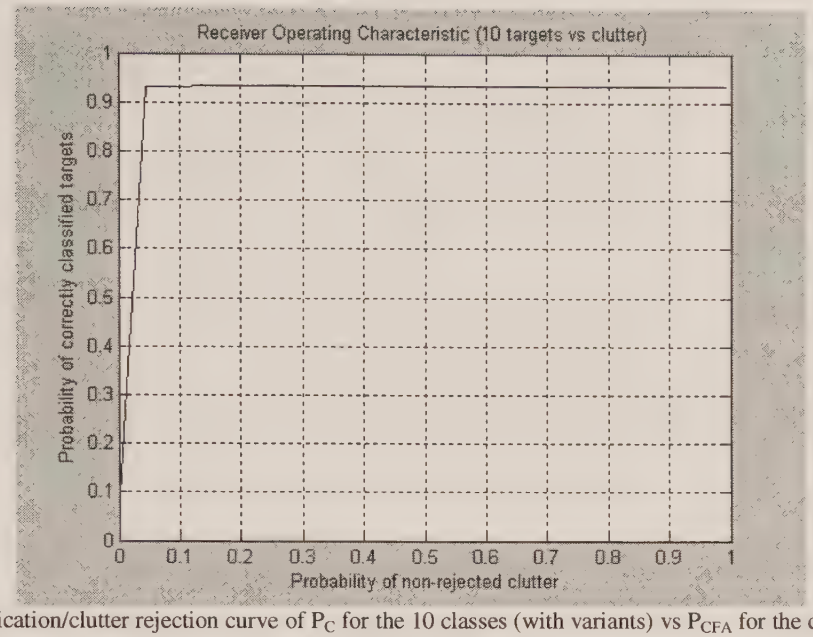


Figure 4.3. ROC classification/clutter rejection curve of P_C for the 10 classes (with variants) vs P_{CFA} for the clutter data using the FST with magnitude FT features

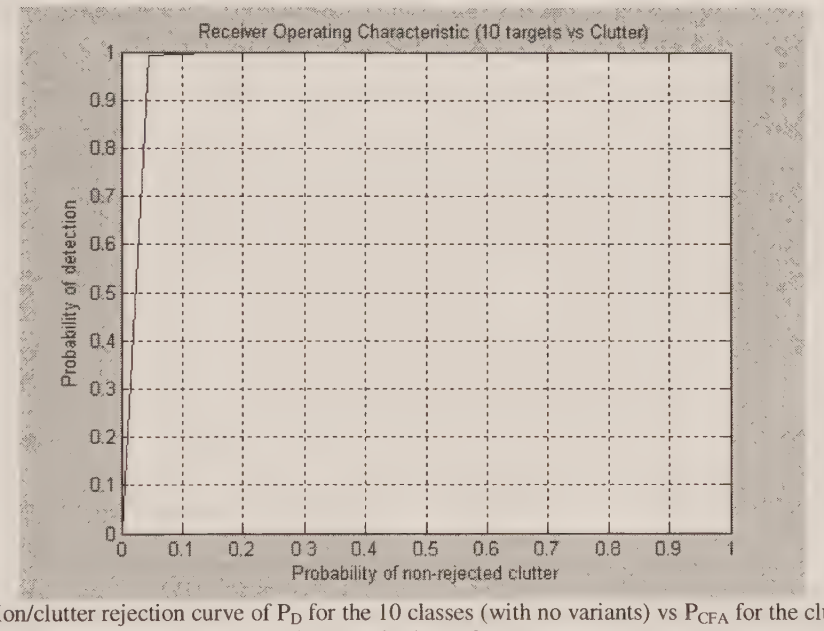


Figure 4.4. ROC detection/clutter rejection curve of P_D for the 10 classes (with no variants) vs P_{CFA} for the clutter data using the FST with magnitude FT features

Table 4.4 presents clutter rejection and target classification scores for 10-class data. It lists P_C and P_{CFA} at T=12.8, a point on the ROC curve. Table 4.4 lists the score at a point on the ROC curve where there is only one false alarm. For the T noted, P_D was high (P_D=94.6%) with no variants; 173 of 2425 targets were missed (P_C=92.9%), but only 1 of 1616 clutter chips (P_{CFA}=0.062%) was not rejected. These scores are for the case of no target variants and only depression angle and aspect view changes. When we considered variants and the same T, P_D=88.5%, 496 of 3203 targets are missed (P_C=84.5%) and the P_{CFA} is the same low score. This is not a very useful operating point because not being able to reject one clutter chip is not of major consequence. An operating point suitable for the application can be chosen from the ROC curve (Figs. 4.3 and 4.4). Clearly, unseen objects (variants of the BMP2 and the T72) cause a significant drop in P_C (~ by 8.4%) and in P_D (~ by 6%). About 323 of 778 or >41% of the variants are missed.

Clutter Rejection	No. of clutter chips tested	No. not rejected as clutter		P_{CFA} (%) T = 12.801	
	1616	1		0.062	
Target classification	No. of targets tested	No. of targets rejected	No. of misclassifications	P_D (%) T=12.801	P_C (%) T=12.801
With variants	3203	367	129	88.54	84.51
Without variants	2425	131	42	94.6	92.9

Table 4.4. Clutter rejection and target classification results, 10 class data, magnitude FT features, FST classifier

In comparison to prior work using distortion invariant filters [6], our scores are much better. They [6] achieved a P_{CFA} of 2.5% at a P_D=90%, for the 10-class data without variants. At another point on our ROC curve, we obtain P_{CFA}=2.5% (as in prior work) at P_C=98.13% and with a much higher P_D=99.7%. P_D is 9.7% more than the P_D obtained in prior work [6].

Acknowledgment

The support of this work by Raytheon Missile Systems is gratefully acknowledged.

5. References

1. T.Ross, S.Worrell, V.Velten, J.Mossing and M.Bryant, "Standard SAR ATR Evaluation Experiments using the MSTAR Public Release Data Set," Proc. SPIE, Vol. 3370, pp.566-573, April, 1998.
2. Abhijit Mahalanobis, Daniel W. Carlson and B. V. K. Vijaya Kumar, "Evaluation of MACH and DCCF correlation filters for SAR ATR using the MSTAR public database," Proc. SPIE, Vol. 3370, pp. 460-468, April, 1998.
3. Abhijit Mahalanobis, Luis Ortiz and B. V. K. Vijaya Kumar, "Performance of the MACH filter and DCCF algorithms on the 10-class public release MSTAR data set," Proc. SPIE, Vol. 3721, pp. 285-291, April, 1999.
4. Anne Marie P. Marinelli, Lance M. Kaplan and Nasser M. Nasrabadi, "SAR ATR using a modified learning vector quantization algorithm," Proc. SPIE, Vol. 3721, pp. 343-354, April, 1999.
5. Michael L. Bryant and Frederick D Garber, "SVM classifier applied to the MSTAR public data set," Proc. SPIE, Vol. 3721, pp 355-260, April, 1999.
6. Rajan Singh and B.V.K. Vijaya Kumar, "Performance of the extended maximum average correlation height (EMACH) filter and the polynomial distance classifier correlation filter (PDCCF) for multi-class SAR detection and classification," Proc. SPIE, Vol. 4727, pp.265-276, Aug, 2002.
7. Abhijit Mahalanobis, B.V.K. Vijaya Kumar, and Robert T. Frankot, "Intraclass and between-class training-image registration for correlation-filter synthesis," Applied Optics, Vol.39, No.17, 10 June 2000, pp. 2918-2924.
8. Rajesh Shenoy and David P. Casasent, "Unconstrained eigen-filter pairs for SAR detection," Proc. SPIE, Vol. 4734, pp. 1-8, Sep, 2002.
9. K. Fugunaka, "Introduction to statistical pattern recognition," New York, NY: Academic Press, 1990.
10. Q. Zhao and J.C.Prinicipe, "Support Vector Machines for SAR Automatic Target Recognition," IEEE Trans. on Aerospace and Electronic Systems, Vol.37, No.2, pp.643-654, 2001.
11. M.L.Bryant, "Target signature manifold methods applied to the MSTAR database: preliminary results," Proc. SPIE, Vol. 4382, pp.389-394, April, 2001.
12. Rajesh Krishna Shenoy, "The Design and use of unconstrained image filters and features for SAR detection and recognition", Ph.D Dissertation in Electrical and Computer Engineering, Carnegie Mellon University, May 2001.
13. Q.H.Pham, A.Ezekiel, M.T.Campbell and M.J.T.Smith, "A New End-to-End SAR ATR System," Proc. SPIE, Vol. 3721, pp. 292-301, April, 1999.
14. Lance M. Kaplan, Romain Murenzi, Edward Asika, Kameswara R. Namuduri, "Effect of signal-to-clutter ratio on template-based ATR," Proc. SPIE, Vol. 3370, pp. 408-419, April, 1998.
15. G.G. Lendaris and G.L.Stanley, "Diffraction Pattern Sampling for Pattern Recognition," Proceedings of the IEEE. Vol.58, No.2, pp.198-216, February 1970.

Face verification and rejection with illumination variations using MINACE filters

Rohit Patnaik and David Casasent

Dept. of ECE, Carnegie Mellon Univ., 5000 Forbes Ave., Pittsburgh, PA 15213

ABSTRACT

A face verification system based on the use of a minimum noise and average correlation energy (MINACE) filter for each person is presented that functions with illumination variations present. A separate filter is used for each person; it is a combination of different training images of only that person. The system is tested using both unregistered and registered images from the CMU Pose, Illumination and Expression (PIE) database. The number of correct (P_C) and the number of false alarm (P_{FA}) scores are compared for the two cases. Rather than using the same parameters for the filter of each person, an automated iterative filter training and synthesis method is used. A validation set of several other faces is used to achieve parameter selection for good rejection performance. For filter-evaluation, all filters are tested against all images, but the same peak threshold is used for each filter to determine verification and rejection.

Keywords: Authentication, biometrics, correlation filters, face recognition, face verification, minimum noise and average correlation energy (MINACE) filter, pattern recognition

1. INTRODUCTION

In face verification, the system must accept or reject the identity claim made by a user based on an image of the user's face. In verification (authentication), the input face is compared only to the claimed reference face and if the match score is above a threshold, the person's identity is verified. In face identification, all filters are checked and the filter with the largest output denotes the class of the input. When rejection of impostor faces (faces not in the database or a face that is in the database but is not the person claimed in verification) is required in identification, the maximum output must be above a threshold or the test input is rejected. In verification, if the filter for the claimed test face output is below threshold, it is rejected. Each of our filters is synthesized from only images of a single person. Our work involves *evaluation* of different face verification systems. In evaluating the performance of our verification system, the output from all filters is analyzed for each test input as a function of the threshold T. Any non-correct filter output $\geq$ T contributes to the false-alarm rate (P_{FA}). As T is varied, ROC (receiver operating characteristic) data (P_C or P_{miss} vs. P_{FA}) is obtained. In the evaluation of an identification system, T is also varied but only the maximum filter output is considered and if it is less than T, the test input is rejected.

Face verification is useful for authentication, security, and e-commerce applications. Our concern in this paper is face recognition in the presence of illumination variations (specifically those in the PIE database). We have interest only in face recognition methods that can handle both illumination and pose variations. In this paper, we evaluate a face verification system based on the minimum noise and average correlation energy (MINACE) filter[1] with illumination variations present, using both unregistered and registered images. We present results using images from the CMU Pose, Illumination, and Expression (PIE) database[2]. Since many of the images in the PIE database are tilted, synthesizing filters from such unregistered images leads to artificially better verification results, since the filters will recognize and discriminate images based on the tilt of the head and not facial structure. MINACE filters emphasize high frequencies.

Prior work on distortion-invariant filters considered face verification[3] (and addressed rejection but only addressed expression differences), and face identification[4] using the illumination differences in the PIE database (but did not address rejection). Both works used the PSR (peak-to-sidelobe-ratio) of the filter output rather than the peak output value for the match score. Fisherfaces[5] have been noted to be of use for illumination variations, but only frontal pose views were considered. Eigenfaces[6] have been shown to handle pose variations (using view-based methods) but they do not perform well on illumination variations[5,7]. Many graphic techniques[8,9,10] have been applied to the PIE and other databases with pose and illumination differences. The light fields method[8] is computationally expensive on-line, and it assumes that a precise pose estimation is available. The morphable 3D face model[9] technique requires 3-D face images from a laser scanner. The last two methods[8,9] have been applied to the PIE database. An illumination cone[10] method has also been used but it requires much storage. The fusion[11] of LDA (linear discriminant analysis) and PCA (principal

Optical Pattern Recognition XV, edited by David P. Casasent, Tien-Hsin Chao, Proc. of SPIE Vol. 5437 (SPIE, Bellingham, WA, 2004) · 0277-786X/04/$15 · doi: 10.1117/12.548082

component analysis) has been applied to face recognition with expression and slight pose differences present (not with illumination differences). Quotient images[12] are illumination-invariant, but the concept cannot be extended to other distortions such as pose variations. In all of this prior work with one exception[3], only verification (authentication) was considered (not evaluation of the system) and thus no rejection tests were performed.

Face identification is generally accepted as being an easier problem that face verification. In this paper, we use an automated iterative filter-synthesis method for the filter of each person; it uses a subset of the training set and incorporates rejection capability into each filter. Section 2 reviews the MINACE filter. Section 3 describes the database. Section 4 describes the filter-synthesis process used to create the filter of each person. Section 5 presents the test results using unregistered and registered images.

2. MINACE FILTER THEORY

Minimum noise and average correlation plane energy (MINACE) distortion-invariant filters have proven to be very attractive for detection and recognition of objects in infrared[1] and SAR[13,14] imagery. Most of the information in this section has been reproduced from Ref. 15. Vectors (matrices) are denoted as lower (upper) case bold letters. All data are in the Fourier Transform (FT) domain, as synthesis of the filter in the frequency domain is easier. The FT of the filter is **h** (all vectors are one-dimensional lexicographically ordered two-dimensional data). The columns of the data matrix **X** are the FTs of training set images (images of a subject under different illuminations). We require the filter to give correlation plane values of one at the origin for each training set image included in the filter synthesis; this *peak constraint* is described by

$$\mathbf{X}^{H}\mathbf{h} = \mathbf{u} = [1,\ldots,1]^{T} \ , \tag{1}$$

where $(\ )^{H}$ denotes conjugate transpose and the elements of the vector **u** are specified correlation peak values for the various training set illumination images. To improve performance, one also requires **h** to minimize a combination of E_S (correlation plane energy due to signal) and E_D (correlation plane energy due to distorted images of a subject). One thus also requires the filter **h** to *minimize the objective function*

$$E = (1 - c)E_S + cE_D \ , \tag{2}$$

where the control parameter c $(0 \leq c \leq 1)$ is chosen to emphasize minimization of E_S or E_D. The correlation plane energy due to all training set images is $\mathbf{h}^{H}\mathbf{S}\mathbf{h} = E_S$, where **S** is a diagonal matrix whose entries are the average power spectrum of the full training set of images at each frequency. The correlation plane energy due to distorted versions of the training set images is similarly $\mathbf{h}^{H}\mathbf{D}\mathbf{h} = E_D$, where the elements of the diagonal matrix **D** follow from some model for distortions; we use a zero-mean white Gaussian noise model with the elements of **D** being all equal to the variance σ^2_N of the noise model. Thus,

$$E = (1 - c)\mathbf{h}^{H}\mathbf{S}\mathbf{h} + c\mathbf{h}^{H}\mathbf{D}\mathbf{h} \ . \tag{3}$$

The Lagrange multiplier solution that satisfies (1) and minimizes (3) is

$$\mathbf{h} = \mathbf{T}^{-1}\mathbf{X}(\mathbf{X}^{H}\mathbf{T}^{-1}\mathbf{X})^{-1}\mathbf{u} \ , \tag{4}$$

where

$$\mathbf{T} = \max[\mathbf{S}_1(u, v), \mathbf{S}_2(u, v),\ldots,c\mathbf{D}] = \text{envelope at each frequency } (u, v) \ , \tag{5}$$

$$c = (\text{dc frequency value of } \mathbf{D}) \,/\, (\text{dc frequency value of } \mathbf{S}) = \sigma^2_N \,/\, \mathbf{S}(0, 0) \ . \tag{6}$$

$\mathbf{S}_i$ is a diagonal matrix whose entries are the power spectrum of training image i; the energy in each training image is normalized to one. The dc value S(0,0) is not normalized to one in this work.

Reducing E_S in (2) subject to the peak constraint in (1) produces delta-function-like correlation peaks; this reduces correlation plane sidelobes and hence localizes the object well in the correlation plane. To achieve this, the filter emphasizes high spatial frequencies in the object data. Hence, minimizing E_S tends to reject imagery not present in the training set; i.e. it minimizes false alarms (P_{FA}), but it can make recognition (P_C) of true-class test set images not present in the training set difficult. Conversely, minimizing E_D emphasizes lower spatial frequencies and thus improves recognition performance (P_C), while making the false alarm performance (P_{FA}) worse. An intermediate value of c is thus preferable. MINACE filter synthesis allows such flexibility and thus offers improved performance over other distortion-invariant filters; it only requires selection of one filter parameter c and is thus easier to design compared to other filters with several parameters to be selected. In conclusion, large (small) values of c improve P_C (P_{FA}). In this paper, the parameter c is adjusted during filter synthesis by varying σ^2_N of the noise model.

3. DATABASE

The complete PIE database[2] contains 41,368 images of 68 people. Each person is imaged under 13 different poses, 43 different illumination conditions, and with 4 different expressions. However, not all of the expression variations and illumination conditions are captured for all the 13 poses. The database we used consisted of 65 people from the illumination subset of the PIE database containing 21 different illumination images per person. This dataset was captured without room lights, and is a harder dataset than one captured with room lights on[4].

Each subject is facing the central (frontal) camera with a neutral expression, and each pose and expression is constant across the 21 illumination differences. The original images are 640 x 486 color images. We converted all of the images to grayscale. We indexed the subjects 1-65 and the various illumination images for each subject were indexed 1-21. Figure 1 shows the original images of subjects 3, 7, 8, 11, 14, 17, 22, 25, and 28 at frontal lighting (Illumination 7). Figure 2 shows the 21 illumination differences for subject 3. The` illumination differences are labeled in the order in which the flashes were indexed in the PIE database. The illuminations vary from extreme left shadow (illumination 1) to extreme right shadow (illumination 16). The other illumination images represent illumination variations between these two extremities; illumination 7 denotes frontal lighting. Figure 2 demonstrates the extreme illumination variations. The wide range of illumination differences is what makes the task of verification challenging. We now describe the process of producing the unregistered and registered images.

Figure 1. Original images of some subjects in the PIE database at frontal lighting (illumination 7)

Figure 2. Illumination differences for subject 3

3.1. Unregistered Images

Based on the locations of features such as the center of the eyes, the tip of the nose, and the center of the mouth, the images were cropped to 128 x 128 pixels. The pixels were then averaged in 3 x 3 overlapping blocks (overlap of one row or column) to reduce the size of the images to 64 x 64 pixels. Figure 3 shows the unregistered images of some of the 65 subjects at frontal lighting (illumination 7). We chose these subjects to demonstrate the different races of people in the database (e.g. subject 1 is Asian, subject 13 is White, subject 21 is Black), and the noticeable tilts present (look at the

eyes and mouth) in some of the images (e.g. subjects 3, 4, 5, 7, 8, 9, 11, 13, 15, 17, 29, 31). To make the task of verification fairer, each of the images should be registered (tilt removed, and eye and nose locations standardized) before creating the face verification system, otherwise recognition will treat tilt as part of facial structure. Note that subject 29 in Fig. 3 has his eyes closed in Fig. 3, this is true for subject 29 in 6 of the 21 illumination images (5, 6, 7, 8, 9, 10).

3.2. Registered Images

The coordinates of four local landmarks (the eyes, mouth, and nose) were located in each image, each image was cropped to 128 x 128 pixels and an affine transform was performed on each image so that the coordinates of the four facial landmarks were the same for all images. All registered images were of size 64 x 64 pixels. Figure 4 shows the registered images of the subjects in Fig. 3. Note that the registered images of different persons in the database look more similar than unregistered images, since image tilt is almost removed in Fig. 4. The affine transformation of images produces distortions in the pixel values. This can, depending on the subject, lead to an increase or a decrease in differences among the various illumination images of the subject. *This is a formidable face recognition problem*; since all faces are registered, discrimination is more difficult; and impostor faces to be rejected are also registered. All unregistered and registered images are normalized to have unit energy before synthesizing the filters and before recognition correlations.

Figure 3. Unregistered images of some of the 65 subjects at frontal lighting (illumination 7)

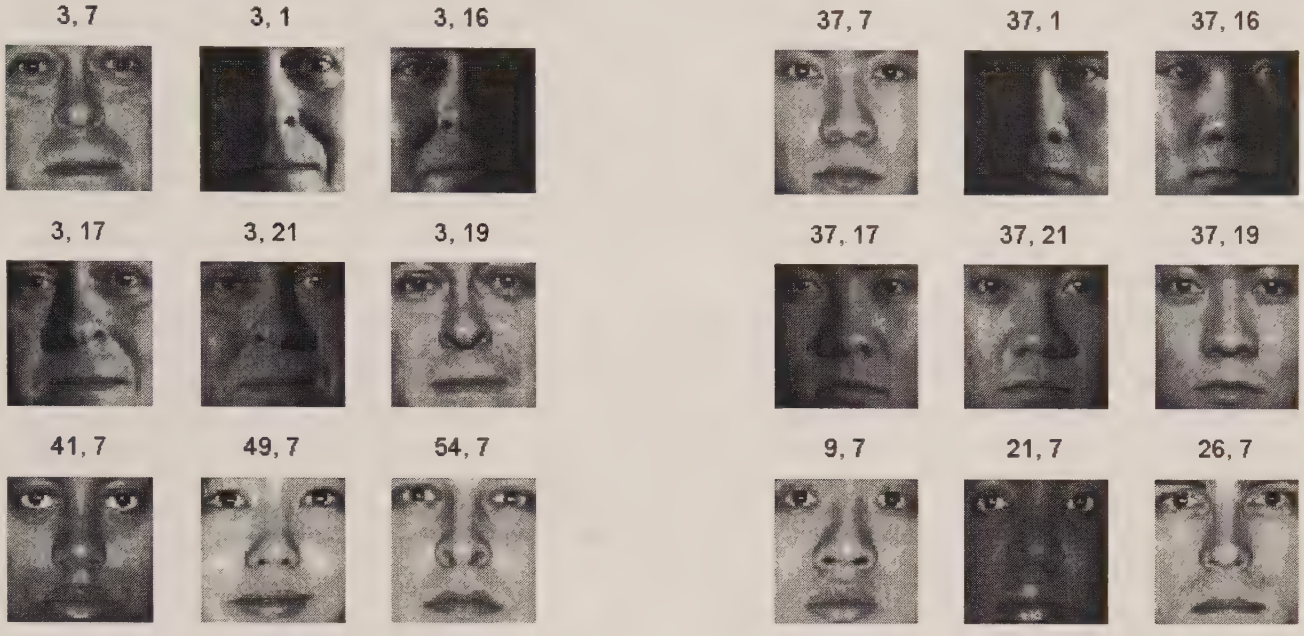
Figure 4. Registered images of some of the 65 subjects at frontal lighting (illumination 7)

4. FILTER-BANK SYNTHESIS

4.1. Training Set

To create the MINACE filter for a subject, a set of training images of the subject under six different illuminations (7, 1, 16, 17, 19, and 21) was input to the filter training program. In addition, three images of other persons in the database were used as a false class validation set to select the filter design parameter c. For the first 33 subjects, the false class validation set consisted of illumination image 7 (frontal lighting) of subjects 41, 49, and 54, i.e., three subjects in the second half of the database. For the last 32 subjects, the false class validation set consisted of illumination image 7 of subjects 9, 21, and 26, i.e., three subjects in the first half of the database. The reason for having two different sets of false class validation images is that if we had used the same false class validation set for the filters of all subjects, then we would have to use a different false class validation set for the three filters of the three persons in the validation set.

Figure 5 shows the registered training and false class validation sets for subjects 3 and 37 respectively. The first two rows correspond to the training set while the third row shows the false class validation set. The pair (i, j) above an image indicates that the image is illumination image j of subject i.



(a) Subject 3

(b) Subject 37

Figure 5. Registered training (rows 1 and 2) and false class validation (row 3) set for (a) subject 3 and (b) subject 37

4.2. MINACE Filter Synthesis

The new automated MINACE filter-synthesis algorithm developed follows:

1. Initialize c to a default value (0.005) for this database. Set the minimum peak constraint for images from the true class (*min_true_pk*) to 0.75 and the maximum peak constraint for the false class validation images (*max_false_pk*) to 0.65. These two peak values were selected for this database from initial trials.
2. Create a filter using one of the images from the training set with this value of c. We arbitrarily used the image with frontal illumination, 7.
3. Correlate this filter with the remaining images in the training set. If any of the correlation peaks are below *min_true_pk*, then add the image with the lowest correlation peak to the filter and synthesize a new filter. Continuing this process ensures that the designed filter will recognize all training set images with correlation peaks $\geq$ some minimum. We note that for training set images included in filter synthesis, the filter (by definition) gives correlation peaks of 1.0 to satisfy Eq. (1).
4. Repeat the correlation in step (3) and continue including more images from the training set as in step (3) until the filter gives correlation peaks at least as great as *min_true_pk* for all images in the training set.
5. If all images in the training set are needed to create the filter, this suggests that the filter does not have good recognition capabilities, i.e., it is not likely to give high correlation peaks for test inputs. In this case, increase the value of c by 0.005 and repeat steps (2) – (4) with the new value of c. Increasing c will make it easier for the filter to recognize different illumination variations.
6. Correlate this filter with the images in the false class validation set. If any of the correlation peaks for the false class validation set images are above *max_false_pk*, then the filter is not discriminative enough. Decrease the value of c by 0.0025 and repeat steps (2) – (5). We decreased c by 0.0025 and not by 0.005 to prevent the filter-synthesis algorithm from getting stuck in a loop.

Once the filter-synthesis algorithm exits step (6), a filter with a good choice of c has been created which should have both good recognition and discrimination capabilities. This concludes the synthesis process for the filter. Steps (2)-(5) ensure good recognition and step (6) ensures good discrimination. This is an automated version of earlier work[10]. Note that for the filters synthesized using both unregistered and registered images, we never had to decrease c in step (6) in the filter synthesis process. This was because of a good initial default choice for c.

4.3. FFT Size

Correlation is performed in the frequency domain using Fast Fourier Transforms (FFTs). In order to perform linear correlation and prevent aliasing, the training and test set images are padded with zeros to 128 x128 pixels before the FFT step, both to create the filter and in performing correlations. One cannot synthesize the filter in the frequency domain as 64 x 64 samples and then zero-pad its image domain representation for performing correlation. Thus, FFTs of 128 x 128 samples are used.

4.4. Filter Data

Figure 6 shows the spatial template (the image-domain point spread function) of the MINACE filters for subjects 3 and 37, synthesized using registered images using the algorithm in Sect. 4.2. Note that the actual spatial template is 128 x 128 pixels but the face shown is only contained in one 64 x 64 pixel quadrant. In Fig. 6, training illumination numbers refers to the illumination variations in the training set that were included in filter synthesis and the order in which the illumination images were chosen in our iterative filter synthesis algorithm. The value of c is also indicated in the figure. As can be seen in Fig. 6, the filter performs edge enhancement and emphasizes key facial structures such as the eyes, nose, and the mouth. This is all done automatically. *This is why a MINACE filter can recognize faces independent of illumination variations.* Use of a large c value (Fig. 6a vs. 6b) implies emphasis of lower spatial frequencies; this is why we see more facial detail in Fig. 6a vs. 6b.

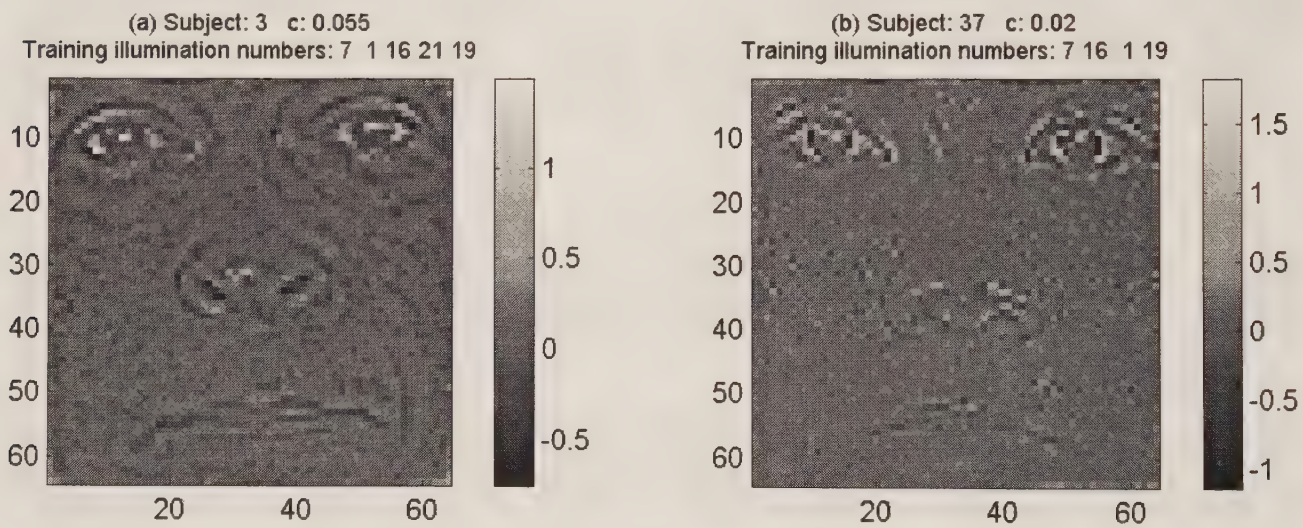


Figure 6. Spatial template of the MINACE filters of (a) subject 3 and (b) subject 37 synthesized using registered images

About 50% of the filters synthesized using unregistered images required five of the possible six illumination images during filter synthesis; 70% of the filters synthesized using registered images required five of the possible six illumination images. To produce registered images, an affine transform is used and this introduces some distortions. This explains why more filters synthesized using registered images required five of the six possible illumination images, as compared to the filters synthesized using unregistered images. For example, subject 6 using registered (unregistered) images and the same $c = 0.005$ value required five (vs. four) training set images (7, 1, 16, 19, 17) vs. (7, 16, 1, 19). All the filters, whether they were synthesized using unregistered or registered images, required frontal lighting (illumination 7, since this was the first illumination chosen), the extreme left shadow (illumination 1), and the extreme right shadow (illumination 16) illumination images in filter synthesis.

In the filter-synthesis step, the frontal illumination image (7) was the first image used in all cases (by definition). In all cases out of 130 (except four), the next two illumination images chosen were the extreme left shadow (illumination 1)

and the extreme right shadow (illumination 16) or illumination 16 followed by illumination 1. All cases included images 7, 1, and 16 as the images used in filter synthesis. This is somewhat expected, but automation is needed, as image 16 was chosen as the fifth image in one case. Then, more images were included from the training set as needed (a maximum of two images from illuminations 17, 19, and 21) to synthesize all filters. For example, the filter of subject 4 synthesized using unregistered images used a c value of 0.005, and illumination images 7, 16, 1, and 17 (in that order), while the filter of subject 57 synthesized using registered images used the same c value of 0.005, and illumination images 7, 1, 16, 21, and 19 (in that order). A few filters required only three training images in filter synthesis. These were the filters of subjects 40, 44, 56, and 63 synthesized using unregistered images, and the filter of subject 63 synthesized using registered images. Thus, most filters (61 of 65 filters synthesized using unregistered images, and 64 of 65 filters synthesized using registered images) required four or five (of the possible six) training images in the filter-synthesis step (five is the maximum number of images we allow in filter synthesis). The filters of subjects 40, 44, and 56 required an additional image when they were synthesized using registered images. This supports our claim that the affine transform introduces distortions in the images.

Nearly all filters used the default value of c (0.005) and thus required no iterations. Thus, the default value of c was obviously chosen after some initial study. We note that four filters synthesized using unregistered images, and five filters synthesized using registered images required iterations and larger c values to recognize the training set. Two cases (filter 37 synthesized using unregistered images, and the filter 65 synthesized using registered images) required only one iteration. The filters of subjects 3, 12 and 21 synthesized using unregistered images required a significantly higher value of c (0.045, 0.04, and 0.095 respectively), while the filters of subjects 1, 3, 10 and 37 synthesized using registered images required significantly higher values of c (0.025, 0.055, 0.045, and 0.02 respectively). We now advance possible reasons for some of the high values of c needed; we do not expect to explain every case.

Larger values of c emphasize lower spatial frequencies and are needed to aid the filter in recognizing distorted images. The high values of c for filters 1 and 37 synthesized using registered images can be attributed to distortions introduced in registered images due to the affine transform. To recognize such distorted transformed images, we require emphasis of lower spatial frequencies (use of a larger c value). This would explain why the filters of subjects 1 and 37 synthesized using unregistered images required low values of c (0.005 and 0.01 respectively), while those created using registered images required higher values of c (0.025 and 0.02 respectively). Subject 3 has curved lines on both sides of his mouth, as is clearly seen in Fig. 5. These curved lines are facial features, but they seem to be unstable with illumination variations, since a large value of c (0.045 and 0.055 for unregistered and registered data respectively) is needed for recognition of true class training faces, i.e., the filter seems to want to omit these lines to aid recognition. With registered images, the transform requires interpolation since the new sample will not be on the Cartesian sample grid; thus, pixel gray levels are interpolated and features can blur (see Fig. 6a, in which only part of the curved line on the right is apparent in the filter).

Subjects 10, 12, and 21 have dark complexions. For all dark subjects, the eye region contains a significant amount of the energy in the image. In general, unregistered images (such as those of subjects 12 and 21) emphasize the eyes more than the corresponding registered images. In these cases, transformation, its interpolation, and the location of the spatial samples seem to enhance other spatial frequencies so that the eyes are less dominant in the transformed and registered images (see Figs. 3 and 4). The effect of illumination differences is worse when the eyes are emphasized (as occurs for dark subjects). Thus, the filters of subjects 12 and 21 synthesized using unregistered images required high values of c; the registered images of these two subjects emphasize the eyes less and thus, the filters created using registered images needed only the default value of c (0.005). The registered images of subject 10 emphasize the eyes more than the unregistered images do (see Figs. 3 and 4, transforming an image can either aid or worsen various images features, depending upon the effect of interpolation and the location of samples). This is why the filter of subject 10 synthesized using registered images required a high value of c, while the filter synthesized using unregistered images required the default value of c (0.005).

5. TEST RESULTS

5.1. Test Set

In verification, one tests if only the correct filter for the stated input class gives an output above a threshold. Our concern is evaluating the system; in this case, we examine both the verification and rejection capabilities of the filters. For the MINACE filter of a particular subject, the true class test set consists of 15 illumination variations of the correct subject that were not used in the training stage, and the false class test set to be rejected consists of all 21 different illuminations

of each of the other 64 subjects. Thus, for a particular filter or subject, the test set consists of 15 true class images to be recognized and 1344 (21x64) false class images to be rejected. This is equivalent to a given image being correlated against all 65 filters, if it is not in the training set of a filter; if an image is in the training set of a filter, it is only correlated against the other 64 filters. Note that in evaluating the system, the filters synthesized using unregistered images are only tested using unregistered images, while those filters synthesized using registered images are only tested using registered images.

5.2. Representative Results using Unregistered versus Registered Images

To illustrate the effect of using unregistered versus registered images in synthesis of the filter and in correlation tests, we show the cross-correlation peak values of the test set versus the MINACE filter for subject 3 using both unregistered and registered images (Figs. 7a and 7b). In Fig. 7, the first horizontal 15 values correspond to the correlation peaks for the 15 true-class test images of subject 3. There is then a spacing of 45 entries on the horizontal axis, after which, the remaining 1344 (21x64) entries are the cross-correlation peak values for the 21 illumination images of the 64 other subjects in the test set. Note that all images are normalized to have unit energy before performing the correlation step.

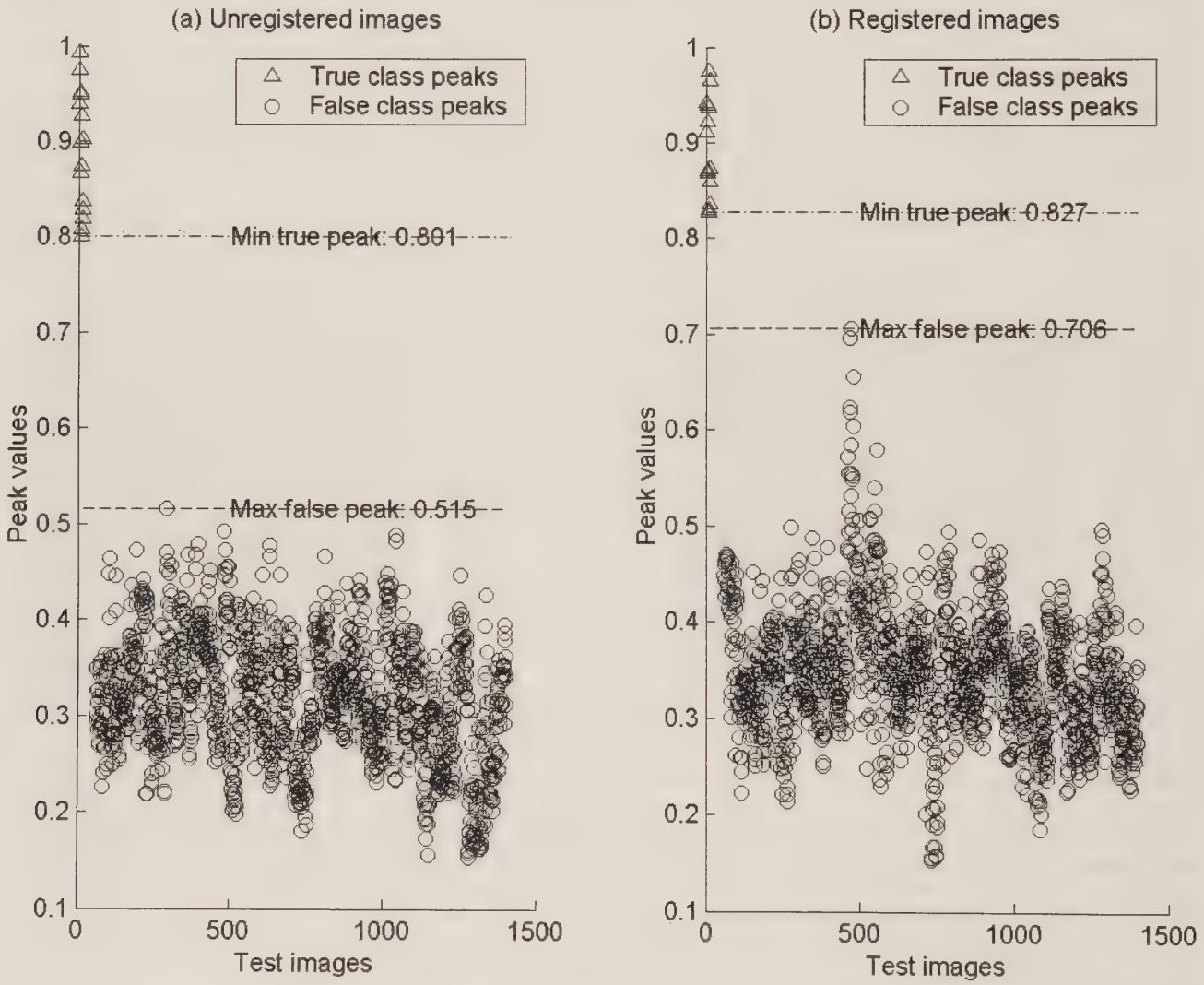


Figure 7. Cross-correlation peak values of the test set vs. the MINACE filter for subject 3 using (a) unregistered images and (b) registered images

From Fig. 7 we see that, for unregistered data, false correlation peaks are smaller (since in registered data, all images are more similar). For unregistered data, all 15 true subject test peaks are above 0.8 and all false peaks are below 0.52.

For registered data, all 15 true subject correlation peaks are above 0.82 and all false correlation peaks are < 0.71 (the false peak values for several subjects are large). For example, nearly all 21 illumination images of subject 21 (illumination images 2, 5, 6, 8, 10-17, 19-21), illumination image 20 of subject 22, illumination images 12 and 13 of subject 24, and illumination images 7, 8 and 12 of subject 25 give false correlation peaks noticeably above 0.5. These large false correlation peak values occur for test images around test image 500 in Fig. 7b (for example, image 25 test results start at test image 543 = 21 illuminations x 23 faces (3 is true class) + 15 true + 45 spaces). We now discuss the reason for some of the large cross-correlation peaks using registered data.

Recall that for registered data, all images have the center of the eyes, the tip of the nose, and the center of the mouth, in the same location; thus all images are more similar. Also, recall that we use normalized images for filter synthesis (Sect. 3.2) and also for performing correlation (Sect. 5.2). Thus, black persons in the database (such as subject 21), will have much more energy located in the eye area compared to the normalized images of other persons in the database. Since all eyes are aligned, black subjects will give stronger cross-correlation peaks than other faces. Since the MINACE filter emphasizes high spatial frequencies, this further emphasizes the eye region and leads to higher correlation peak values for such false class test set images. Recall that only use of registered images gives fair results. We note that for subject 3, all true cross-correlations are noticeably above the largest false correlation. Thus, selection of the proper threshold T allows perfect results. We now present summary data for all 65 faces.

5.3. Performance Results

We evaluate the performance of the full system as the correlation peak threshold T is varied. For the filter of a particular subject, if an input image produces a correlation peak at least as great as the peak threshold, the image is classified as belonging to that subject (it is recognized), otherwise it is classified as not belonging to that subject (it is rejected). The number of true correct samples for a particular filter is equal to the total number of true class images in the test set that the filter recognizes; the number of misses for a particular filter is equal to the number of true-class images that the filter rejects. Note that the sum of the number of true correct samples and the number of misses equals 15 (the number of true-class images in the test set). The number of false alarms is equal to the total number of false images in the test set that the filter does not reject (the filter erroneously recognizes them as true-class images). The maximum number of false alarms for any filter is 1344 = 21 x 64 (the number of false-class images in the test set per filter).

The average number of misses per filter (P_{miss}) for a system is found by dividing the total number of misses for all the filters by the total number of filters, 65. The average number of false alarms per filter (P_{FA}) is the ratio of the total number of false alarms to the total number of filters, 65. The receiver operator characteristic (ROC) curve is the plot of P_{miss} versus P_{FA} for different values of the peak threshold T. From the ROC curve, one can choose an appropriate operating point. For both systems (unregistered and registered) we varied the peak threshold from 0.8 to 0.6 in decrements of 0.005. For unregistered data, as the peak threshold was varied from 0.8 to 0.63, P_{FA} was zero, while P_{miss} decreased from 1.71 to 0.015 misses per filter (averaged); this corresponds to percentages P_{miss} = 1.71/15 = 11.4% and 0.015/15 = 0.1%. For peak threshold values between 0.625 and 0.61, P_{FA} was still zero while P_{miss} was also zero (perfect performance). Decreasing T further resulted in one false alarm in total for all 65 filters (a P_{FA} of 0.0154 per filter) for $0.605 \geq T \geq 0.6$, and P_{miss} = 0. Thus, the best system for unregistered data (for peak thresholds in the range 0.61-0.625) has 15 correct recognitions for each filter (P_C = 100%), and zero false alarms for each filter (P_{FA} = 0%).

Figure 8 shows the ROC curve for registered data for correlation peak thresholds between 0.8 and 0.665. Some of the data points are labeled by their corresponding correlation peak threshold. A value of 0.0154 (1/65) along either axis indicates a total of one false acceptance error or one miss for all 65 filters. Figure 8 also shows the equal error rate (EER) line (a line joining the set of points for which P_{miss} equals P_{FA}). This line intersects the ROC curve at the point corresponding to a peak threshold of 0.695, that is, for registered data, a peak threshold of 0.695 leads to equal P_{miss} and P_{FA} of 0.108 (or a total of 0.108 x 65 = 7 false alarms errors and miss errors in all 65 filters). We analyzed the errors at this T = 0.695 operating point. As expected, several filters have most errors. Filters 3, 14, and 15 have two false alarms each and filter 65 has one false alarm. All false alarms are because of various illumination images of subject 21. As indicated in our discussion in Sect. 5.2, correlation with images of black persons in the database (such as subject 21) lead to high correlation peak values for the registered images. Filters 2, 15, 27, 36, and 40 have one miss each (the misses occur for illumination images 15, 5, 5, 2, and 15 respectively), and filter 23 has two misses (illumination images 3 and 4). The total number of true-class inputs to a system is 975 (15x65), while the total number of false-class inputs is 87360 (21x64x65). Thus, for a peak threshold of 0.695, P_C for registered data is 99.28% (1358/1365), while P_{FA} is 0.008% (7/87360). Note that a P_{miss} of 0.0733% indicates one missed verification (in total), while a P_{FA} of 0.0011% indicates one false alarm (in total).

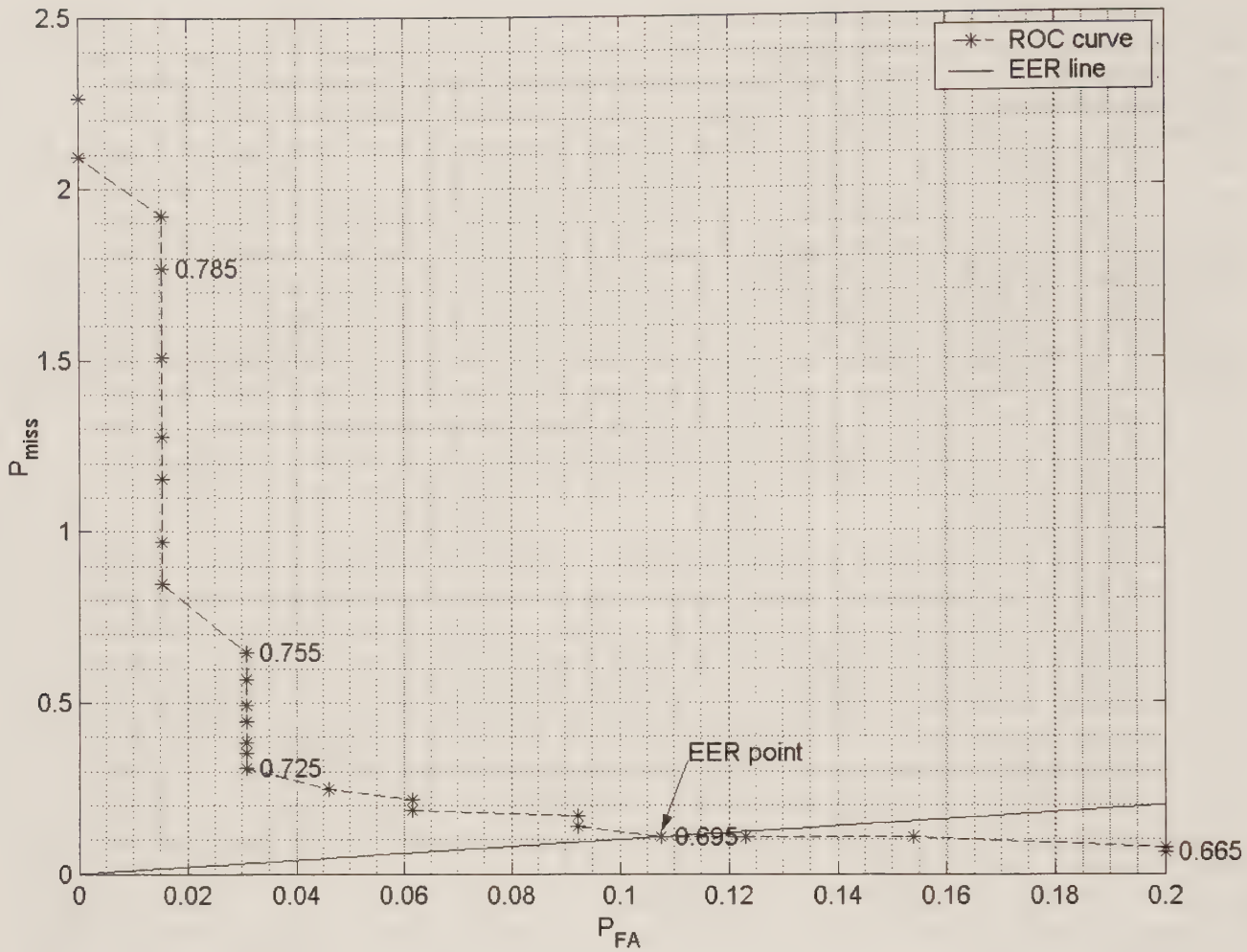


Figure 8. ROC curve for registered data: average number of misses (P_{miss}) per filter (out of 15) vs. average number of false alarms (P_{FA}) per filter (out of 21 x 64 = 1344). Selected correlation peak values are noted

6. SUMMARY AND CONCLUSION

An automated iterative MINACE filter-synthesis method for creating a face filter of each person invariant to illumination differences has been presented. A face verification system using this method was evaluated using both unregistered and registered images. The automated filter-synthesis method uses a subset of the training images and incorporates the rejection ability of each filter in the filter-synthesis step; it is shown to provide high P_C and low P_{FA}.

From an analysis of errors, we found that the eye regions in the images of people with dark complexions contain a significant amount of energy. The more the eye area is emphasized, the worse the effect of illumination differences becomes and a higher value of c is required during filter synthesis to emphasize lower spatial frequencies. It was also observed and explained that for registered data, the images of a few black persons in the database produced high false correlation peaks; this resulted in false alarms for registered data in our test results.

The best system for unregistered data (P_C = 100%, P_{FA} = 0%) is clearly better than the best system for registered data (P_C = 99.28%, P_{FA} = 0.008%). However, registering images introduce certain distortions in the images and results in all images of different persons being more similar, thus producing a harder verification problem.

ACKNOWLEDGMENTS

The authors thank Chao Yuan of Carnegie Mellon University for providing the coordinates of the key facial features and for providing the affine transform code used for registering images.

REFERENCES

1. G. Ravichandran and D. Casasent, "Minimum noise and correlation energy filters," *Appl. Opt.* **31**, pp. 1823-1833, April 1992.
2. T. Sim, S. Baker, and M. Bsat, "The CMU Pose, Illumination, and Expression (PIE) Database", *Proc. Of the IEEE International Conference on Automatic Face and Gesture Recognition*, May 2002.
3. M. Savvides, B.V.K. Vijaya Kumar and P.K. Khosla, "Face verification using correlation filters," *Proc. Of the Third IEEE Workshop on Automatic Identification Advanced Technologies*, pp. 56-61, Tarrytown, NY, March 2002.
4. M. Savvides, B.V.K. Vijaya Kumar, "Illumination-tolerant face recognition using advanced correlation filters," submitted to *Computer Vision and Pattern Recognition*, 2003.
5. P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **19**(7), pp. 711-720, July 1997.
6. A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular Eigenspaces for face recognition," *Computer Vision and Pattern Recognition*, pp. 84-91, June 1994.
7. D. Casasent and C. Yuan, "Face recognition with pose and illumination variations using new SVRDM support vector machine," *Proc. SPIE* **5437**, paper 1, April 2004.
8. R. Gross, I. Matthews, and S. Baker, "Appearance-based face recognition and Light-fields," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **26**(4), April 2004.
9. V. Blanz, S. Romdhani, and T. Vetter, "Face identification across different poses and illuminations with a 3D morphable model," *Proc. Of the Fifth IEEE Conference on Automatic Face and Gesture Recognition*, pp. 202-207, May 2002.
10. A.S. Georghiades, P.N. Belhumeur, D.J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **23**(6), pp. 643-660, June 2001.
11. G. L. Marcialis and F. Roli, "Fusion of LDA and PCA for face verification," in *Biometric Authentication*, vol. LNCS **2359**, pp. 30–37, June 2002.
12. A. Shashua and T. Riklin-Raviv, "The quotient image: Class-based re-rendering and recognition with varying illuminations," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **23**(2), pp. 129-139, February 2001.
13. D. Casasent and R. Shenoy, "Synthetic aperture radar detection and clutter rejection MINACE filters," *Pattern Recognition* **30**(1), pp. 151-161, 1997.
14. D. Casasent and S. Ashizawa, "Synthetic aperture radar detection, recognition, and clutter rejection with new minimum noise and correlation energy filters," *Optical Engineering* **36**, pp. 2729-2736, October 1997.
15. D. Casasent, K, Ippolito, and J. Verley, "Distortion-invariant FOPEN detection filter improvements," in *Algorithms for Synthetic Aperture Radar Imagery VI*, E. Zelnio, ed., *Proc. SPIE* **3721**, pp 261-271, 1999.

Theoretical modeling of the operational mechanism of a photoconductive MEMS spatial light modulator under AC and DC bias

J. Khoury and C. L. Woods
Air Force Research Laboratory / SNHC, Hanscom Air Force Base, MA 01731

B. Haji-Saeed, S. K. Sengupta, D. B. Megherbi and W. Goodhue
Physics and Electrical and Computer Engineering Departments,
University of Massachusetts Lowell, Lowell MA 01854

J. Kierstead
Solid State Scientific Corporation, Hollis NH 03049

ABSTRACT

This paper analyzes the operation of a new optically addressed deformable mirror device for applications in adaptive optics and optical signal processing. Device operation utilizes a pixellated metallized polymeric membrane mirror supported above an optically addressed photoconductive substrate. A grid of patterned photoresist supports the metallized membrane. A conductive ZnO layer is placed on the backside of the substrate. The device operates as an impedance distribution between two cascaded impedances between the deformable membrane and substrate and the substrate and back electrode. We develop a theoretical model to analyze the deformation as a function of the light intensity and electrical drive.

Keywords: MEMS, Deformable Mirrors, Spatial Light Modulator.

1.INTRODUCTION

Deformable mirror spatial light modulators are the key enabling components of many optical signal processing systems including optical correlators and adaptive optics systems for atmospheric turbulence correction. Such mirrors may be formed of many independently actuated rigid sections (segmented mirrors) or with large area reflective flexible membranes (membrane mirrors) with suitable actuators. Currently, there are three main technologies for addressing the membrane mirror light modulation (MLM) [1]. This includes electron beam addressing [2,3], optical addressing [4-7] and electrical addressing via integrated circuits [8-10]. Membrane mirror spatial light modulators have been used as large-screen display systems [2], infrared scene projectors [3] and light shutters [1]. In the first two designs there is a charge-transfer-plate (CTP), an insulating matrix surrounding an array of conductors, which would be addressed by the electron-beam in the back and the reflective membrane is suitably supported on an etched grid on the front side. In a membrane mirror light shutter, the entire membrane itself acts as one of the electrodes with a small number of discrete hard-wired counter electrodes electrostatically addressing the entire device. While some deformable mirrors of necessity utilize very large apertures, the current trend is towards miniaturization wherever system requirements will allow. Some approaches in adaptive optics related to VLSI hybrid wavefront reconstructor chips and micro-electromechanical systems (MEMS) micro-mirror array for reducing the size and weight of satellite systems or wavefront correction using MEMS Micro-mirrors, also have been done recently [11]. A sophisticated and complicated four-layer planarized SUMMiT4 (Sandia ultraplanar multi-level MEMS technology) process has been used in this process.

This paper analyzes the operation of a new optically addressed deformable mirror device for applications in adaptive optics and optical signal processing. The device consists of a pixellated metallized membrane mirror suspended over an optically addressed photoconductive substrate such as GaAs or InP. A grid of insulating material such as photoresist is used to support the suspended membrane.

A DC or AC bias voltage is applied between the membrane and the transparent ZnO back electrode. The mechanism behind the operation of this device is as follow: The illumination of light on the back of the device increases the

Optical Pattern Recognition XV, edited by David P. Casasent, Tien-Hsin Chao, Proc. of SPIE Vol. 5437 (SPIE, Bellingham, WA, 2004) · 0277-786X/04/$15 · doi: 10.1117/12.548074

conductivity of the photoconductive layer, which leads to a redistribution of the effective electrostatic field between the suspended membrane and the front-side of the semiconductor. This increases the deflection of the membrane in the area of illumination.

We developed two different theoretical models to analyze membrane deformation as a function of the light intensity, for DC and AC bias conditions. The device can be analyzed as two separate impedances, one between the deformable membrane and substrate, the other between the substrate and back electrode. In the DC bias case, we consider our two impedances as two capacitors in the absence of light, and as a resistor and a capacitor for the illuminated device. We derive an expression for the mirror deflection, which is independent of light intensity and quadratic in the amplitude of the DC bias. For the AC bias case, our analytical expression for the membrane deflection has been developed by the assumption of a cascade of a capacitor and a resistor whose resistance is dependent on the light intensity. The results show that the mirror deflection drops as a function of the modulation frequency, saturates as a function of the light intensity and increases quadratically as a function of the amplitude of the AC drive.

2. DEVICE ARCHITECTURE

Figure 1 shows the architecture of our device. It is fabricated on a semi-insulating GaAs or InP wafer. A transparent conducting ZnO layer is deposited on the backside. A grid of photolithographically patterned developed photoresist on the front-side serves as the membrane support. The deformable membrane is thermally bonded to the support to complete the device. The device operates as follows: a DC bias voltage or a high frequency AC voltage is applied between the metallized membrane and the back contact. When a light is shone on the back of the device, the conductivity of the GaAs layer increases. This leads to a higher effective electrostatic drive between the suspended membrane and the front-side of the semiconductor, causing increased membrane deflection in the illuminated areas.

3. THEORETICAL MODEL

We provide here our simplified basic analysis of device operation. A more detailed treatment will be published elsewhere.

3.1 Theory for DC bias operation

A DC voltage is applied to the MEMS device. The equivalent circuit of each pixel of the device may be described as two capacitors C_1 and C_2 in series, where they are the capacitances associated with the photoconductive substrate and the membrane-substrate interface, respectively. The effect is illustrated in figure 2. By the assumption of parabolicity of the deflection of membranes, in the first order approximation, the capacitance of the membrane C_2 is derived as:

$$C_2 \cong \frac{\pi\varepsilon_0 r_1^2}{s} + \frac{\pi\varepsilon_0 r_1^2 h}{2s^2} \qquad (1)$$

where ε_0 is the permitivity of free space, $2r_1$ is the size of one square of the pixel pattern (the experimental devices utilize square pixels, while we assume circular geometry pixels for the analysis itself), s is the depth of the well and h is the deflection of membrane [2,3], which may be modeled as:

$$h = \frac{\varepsilon_0 r_1^2 V^2}{8Ts^2} \qquad (2)$$

where T is the tension on the membrane. The first term in equation (1) is for plane parallel-plate geometry, while the second term adds the additional capacitance due to the curvature of the top plate of the capacitor. The total capacitance of the device can be written as:

$$C = \frac{C_1 C_2}{C_1 + C_2} \qquad (3)$$

where C_1 and C_2, as before, are the capacitances of the photoconductive material and membrane, respectively. C_1 and C_2 form a capacitive voltage divider, so the voltage across the membrane is given by:

$$V_2 = \frac{C_1 V}{C_1 + C_2} \qquad (4)$$

By substituting equation 1 in equation 4 and neglecting the higher order terms we can solve Eq. 4 to find V_2, the voltage across the membrane, as follows:

$$V_2 \cong \frac{C_1 V s}{C_1 s + \pi \varepsilon_0 r_1^2} \quad (5)$$

Substituting equation 5 in equation 2 gives us the deflection of the membrane before light illumination on the back as:

$$h = \frac{\varepsilon_0 r_1^2 C_1^2 V^2}{8T(C_1 s + \varepsilon_0 r_1^2 \pi)^2} \quad (6)$$

After light illumination the device no longer acts like two capacitors in series; it acts as a series RC circuit, in which the membrane remains as a capacitor but the photoconductor changes to a resistor. That is, the device equivalent circuit changes over from a C-C (capacitor-capacitor) circuit to a RC circuit. This is illustrated in figure 3. In this case, in the steady state, the applied voltage drop is just across the capacitor (membrane); therefore the deflection of the membrane changes to:

$$h' = \frac{\varepsilon_0 r_1^2 V^2}{8Ts^2} \quad (7)$$

And the change in deflection upon illumination can be written as:

$$\Delta h = h - h' = \frac{\varepsilon_0 r_1^2 V^2}{8Ts^2}\left(\frac{C_1^2}{(C_1 + C_0)^2} - 1\right) \quad (8)$$

where C_0 is the capacitance of the membrane when there is no deflection (h=0). From this equation, under DC bias operating conditions $C_0 < C_1$, therefore there is only a small deflection of the membrane. It is important to note that the deflection is independent of light intensity – only two states of deflection corresponding to the light being on or off are possible. That is, binary operation is achieved with optical addressing.

3.2 Theory for AC bias operation

By applying high frequency AC voltage on the MEMS, two impedances Z_1 and Z_2, which are in series, will be created between deformable membrane substrate, substrate and electrode respectively. Figure 4 shows the circuit. Before the back illumination on MEMS the voltage across the membrane is:

$$V_1 = V\frac{Z_1}{Z_1 + Z_2} \quad (9)$$

where V is the applied voltage across the MEMS and $Z_2 = \rho\frac{L}{A} = \frac{L}{q\mu_n nA}$.

q is the electronic charge, μ_n is the carrier mobility, n is the carrier concentration under dark illumination conditions, L is substrate thickness and A is the area of each pixel.

After shining light Z_2 will be changed to:

$$Z_2' = \frac{L}{q\mu_n(n + \Delta n)A}$$

where Δn is the increase in carrier concentration due to illumination. By substituting Z_2' and $Z_1 = \frac{1}{j\omega c_1}$, where c_1 is the capacitance of the membrane defined by equation 1 and ω is the frequency, in equation 9, the voltage around the MEMS after shining light is:

$$V_1' = \frac{V}{1 + \frac{Lj\omega c_1}{q\mu_n(n + \Delta n)A}} \quad (10)$$

Substituting equations 1 and 2 in equation 10, gives us an approximated

$$V_1' : |V_1'|^2 \cong \frac{q^2\mu_n^2(n + \Delta n)^2 A^2 V^2}{q^2\mu_n^2(n + \Delta n)^2 A^2 + L^2\omega^2\left(\frac{\pi\varepsilon_0 r_1^2}{s}\right)^2}$$

This equation has been derived by assuming of $\varepsilon_0 << 0$.
Now this equation can be used in equation 3 for finding the deflection of the membrane, which is:

$$\Delta h = \frac{\varepsilon_0^{\,3} r_1^{\,6} q^2 \mu_n^{\,2} A^2 V^2 L^2 \omega^2 \pi^2 (2n\Delta n + \Delta n^2)}{[s^2 q^2 \mu_n^{\,2} (n+\Delta n)^2 A^2 + L^2 \omega^2 \pi^2 \varepsilon_0^{\,2} r_1^{\,4}][s^2 q^2 \mu_n^{\,2} n^2 A^2 + L^2 \omega^2 \pi^2 \varepsilon_0^{\,2} r_1^{\,4}]}$$

According to this formula, for low light intensities, the deflection is quadratic in the total light intensity incident on the backside of the device, and saturates as a function of the light intensity. Furthermore, the deflection drops as the modulation frequency increases. This can be explained as follow: when the modulation frequency is very high the impedance between the membrane and the substrate becomes very low, hence all the voltage drops across the substrate, and almost no voltage drops across the membrane and substrate. This leads to a very small deflection in the membrane.

5.CONCLUSIONS

In this paper we propose an analytical model for an optically addressed electrically driven deformable mirror spatial light modulator. We develop a theoretical model to show the membrane deformation as a function of the light intensity. The device consists of a pixellated metallized membrane mirror suspended over an optically controlled photoconductive substrate. It operates as follows: a bias voltage is applied between the membrane and the back contact transparent electrode. When a light is shone on the back of the device, the conductivity of the photoconductive layer increases. This leads to a higher effective electrostatic drive between the suspended membrane and the front-side of the semiconductor, causing increased membrane deflection in the illuminated areas.
The device was analyzed as an impedance distribution consisting of the impedance of the deformable membrane substrate and the impedance between the substrate and back electrode. For the DC bias theory, we assumed that our two impedances are a cascade of two capacitors in the dark and of a capacitor and a resistor, when illuminated. Based on this assumption, we derive an analytical expression for the mirror deflection as a function of the light intensity. In our results, we found that the mirror deflection is independent of the light intensity that is shone in the back of the device, and is quadratic with the DC bias. For the AC bias case, we developed a theoretical expression for the membrane deflection under the conditions where the AC modulation frequency is much higher than the frequency response of the deformable membrane. In this case the equivalent electronic circuit of the device was assumed to be a cascade of a capacitor and a resistor whose resistance is dependent on the light intensity. In our results we found that the mirror deflection drops as a function of the modulation frequency, saturates as a function of the light intensity and increases quadratically as a function of the AC amplitude modulation.

REFERENCES

1. C. Warde, J. McCann, J.MV.Shrauger, H. Ieong, A.Ersen, X-Y Wang and J. Hubbard "Membrane Mirror Light Modulator Technology", Proceedings of SPIE 191,Vol.3951 (2000).
2. C. Warde, J. E. Hubbard, Jr., G. Genetti, L. Lerman, W.Loizides "Membrane Mirror Light Valve for High-Definition Projection Display", Proceedings of SPIE 270,Vol.1910 (1993).
3. C. Warde, J. E. Hubbard, Jr." Membrane -Mirror- Light -Valve-Based Infrared Scene Projector" Proceedings of SPIE 544,Vol.2223 (1994).
4. L.E. Somers," The photoemitter membrane light modulator image transducer," in Advances in Electronics and Electron Physics, Vol.33A (Academic Press, New York, 1972), 493.
5. F. Reizman,"An Optical Spatial Phase Modulator Array Activated by Optical Signals," in Proceedings of the 1969 Electro-optical Systems Design Conference, pp225-230.
6. A.D.Fisher, L.C. Ling, J.N. Lee and R.C. Fukuda,"Photoemitter Membrane Light Modulator," Opt. Eng. 25, 261-268,1986.
7. D.R. Pape,"Optically Addressed Membrane Spatial Light Modulator," Opt.Eng. 24, 107-110, 1985.
8. K. Preston, Jr.,"An Array Optical Spatial Phase Modulator," Proceedings of the IEEE International Solid State Circuits Conference, NewYork, 1968, p.100.
9. D.R. Pape and L.J. Hornbeck,"Characteristics of the Deformable Mirror Device for Optical Information Processing," Opt.Eng. 22, 675-681,1983.
10. L.J. Hornbeck,"Deformable-Mirror Spatial Light Modulators," SPIE Critical Reviews 1150, 86-102, 1989.
11. Natalie Clark, "A Silicon Eye using MEMS Micromirrors," MRS BULLETIN, 320-324, April 2001.

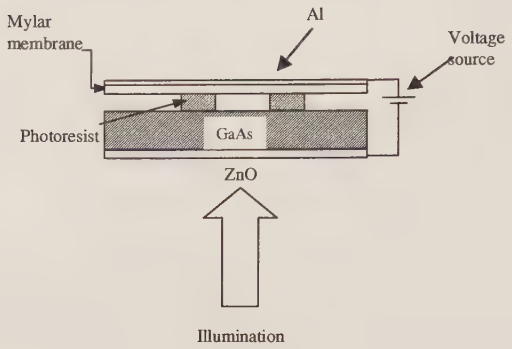


Fig.1. Photoconductive Optically Controlled Spatial Light Modulator

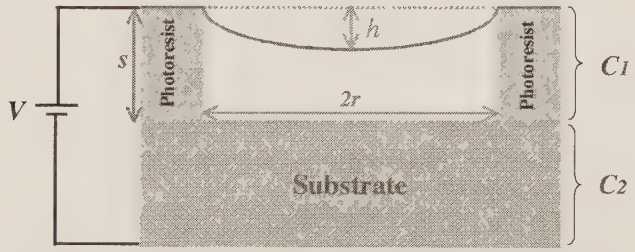


Fig.2. The membrane driven by Dc voltage before the light illumination

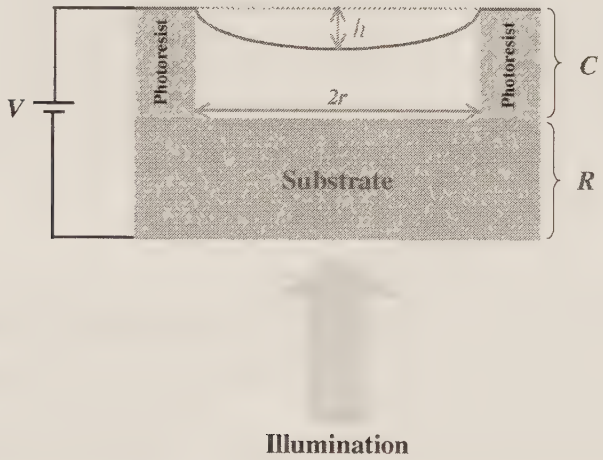


Fig.3. The membrane driven by Dc voltage after the light illumination

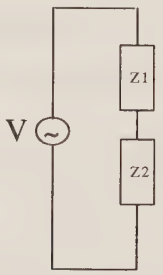


Fig.4. The MEMS device driven by AC voltage

A Real-time Optical Automatic Target Recognition System

Huainxin Chen[a, b], Jianshe Nan[a], Xiaosun Li[b], and Honggang Wei[b]

[a]Southwest China Institute of Electronic Technology, Chengdu, P.R. China;
Phone: + 86 +28-87555717, Fax:+ 86 +28-87555519,E-mail: chenhuaixin@sina.com

[b]School of Electronic and Information, Sichuan University,
Chengdu, P.R. China

ABSTRACT

Automatic target recognition (ATR) technique has been applied in both civil and military. In this paper, we present a new optical pattern recognition system for target recognition. This system includes synthetic discriminate function (SDF) based practical optimized filters for the 3-D targets, the Reference Filter Libs for high correlation SNR, the mapping between the input (object regions) and the output (correlation peaks), and neural networks (ANN) for final decision making. The Real-time optical target recognition is realized by temporal multiplexing technique with electronic addressing spatial light modulator. The experiment results show that the proposed OPR system is efficient and reliable.

Keywords: Optical Pattern Recognition; Synthetic Discriminate Function; Reference Filter Labs; Temporal Multiplexing Technique; Neural Networks.

1. INTRUDCTION

Automatic target recognition (ATR) has general applications in both civil and military, such as recognizing and tracking aircrafts in battle fields, and managing vehicles in modern city traffics. Optical pattern recognition (OPR) techniques are promising because of their merits of high-speed parallel operation, high spatial resolution, flexibility and accuracy.

For the use of OPR systems in practical applications, it is important to develop techniques to handle 3-D object distortion problems (such as the changes of an object in scale, orientation and azimuth). One efficient method is based on Synthetic Discriminate Function (SDF)[1], which uses training image sets to represent 3-D object distortions. Usually, the SDF filter needs selected training sample set covering most of the distortion range, but the number of samples in the set should be as small as possible. Therefore, it is necessary to design the optimized Reference Filter Libs (RFL) to provide high correlation signal noise ratios(SNR) and process data in real-time.

To ensure a high recognition rate and reliability in optical pattern recognition systems, choice of the method of the correlation decision making method is very important. For SDF Filters, the correlation decision making is based on the intensity of correlated peak and a suitable threshold. The use of threshold, however, can

Optical Pattern Recognition XV, edited by David P. Casasent, Tien-Hsin Chao, Proc. of SPIE Vol. 5437 (SPIE, Bellingham, WA, 2004) · 0277-786X/04/$15 · doi: 10.1117/12.541420

only eliminate low level noises in the noisy scenes. It can't reject high level noise sources or false objects while identifying all true auto-correlations correctly. Booth proposed[2] the automatic post processing method of correlation plane imagery using neural networks (ANN) to improve the recognition rate and reliability. ANN operates in a robust, fault tolerant manner and can respond to environment changes through the use of well defined learning rules. In this paper, we present a ANN based post processing for the final decision making. Our method includes two steps: first, employing the mapping relation between the input plane (images in scenes) and output plane (correlation peaks) to find the small ROI with target; second, recognizing the targets again based on ANN to improve the system reliability.

The Real-time hybrid optical digital pattern recognition system is performed with electronic addressing liquid crystal spatial light modulator (LCSLM) for optical correlation, a temporal-spatial multiplexing technique for filtering and ANN for post processing.

2. OPR ARCHITECTURE AND METHODS

2.1 Real-time optical target recognition system

The real-time optical target recognition system proposed in this paper is showed in figure 1. The target images are grabbed by CCD1 and preprocessed by digital edge operators in the computer. Target images and SDF filters are fed to SLM1 and SLM2 respectively. The SDF filter is generated off-line with a group of training images set that covers limit distortion range of 3-D object. The correlation output are detected by CCD2; the optical correlation operation is performed by optical filtering based on renewable SDF filter in real-time; the correlating peak signals are processed by threshold technique, and finally targets are recognized using ANN.

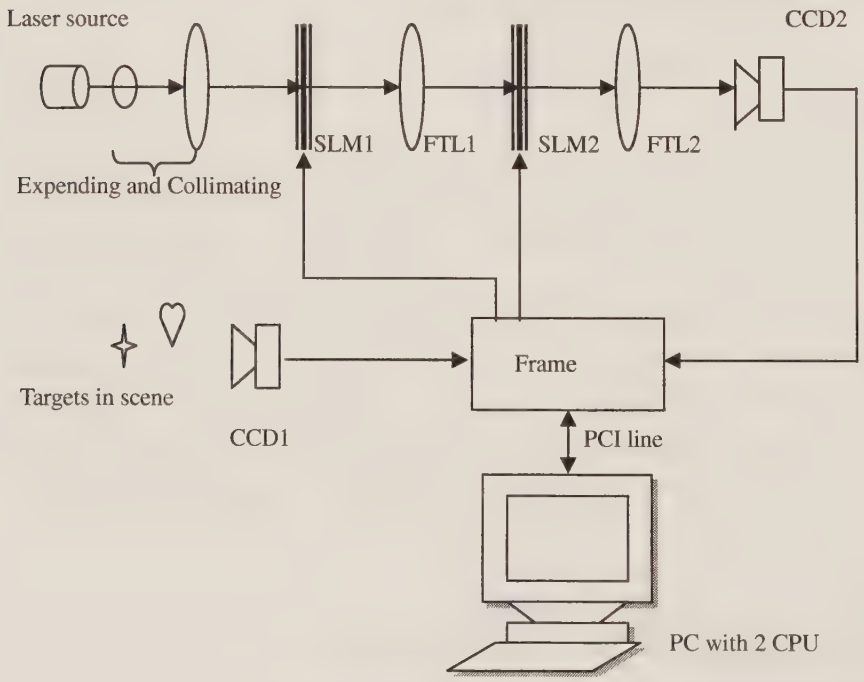


Fig.1 A real-time optical target recognition system architecture

2.2 Optimized SDF Reference Filter Libs

Synthetic Discriminate Function (SDF) filter is a distortion-invariant composite filters, using training sample images to represent statistically a 3-D object distortion vision. There are two types of composite filter: the Maximum Average Correlation Height (MACE)[3] and the Minimum Average Correlation Energy (MACH)[4]. The modified equation correlation peaks (ECP) SDF filters are used in our OPR system as they can be easily implemented with optical hardware.

Usually, the researches on distortion-invariant recognition focus on developing the algorithm. It is also important to study the tolerance range of target distortions in order to optimize SDF filters in optical hardware implementation. To realize the 3-D object distortion recognition, a large set of training images are required. This will arise two problems: one is that the SNR of correlation peaks decreases as the number of training images increases; the other is the difficulties to implement by any optical-electric devices because of a large dynamic range of amplitude and phase encoded in the filter. Therefore, one of the challenge problems in SDF Filter is to determine the best combination of objects for one SDF filter, that is, to design the optimized reference filter libs.

To reduce the large gray dynamic range of SDF filter and to improve the SNR of correlation output, we separate training images into several groups. Each group that consists of a training subset with the coverage of a limited distorted range corresponds to one distortion-invariant SDF filter. In the real-time hardware optical implementation for SDF filter target recognition, there are two different approaches: one is the spatial multiplexing technique, which requires multi-optical channels and thus increases the complexity of OPR system; the other is the temporal multiplexing technique, which requires the SDF filter. The temporal multiplexing technique is attractive as it is high rate renewable. Especially, new commercially available electric addressing LCSLM has shown high renewal performance. Therefore, we employ the temporal multiplied technology in the OPR system, in which a set of SDF filters are renewed by electric addressing LCSLM in real-time to perform recognition for 3-D distortion targets.

For 3-D distortion-invariant recognition problems, understanding the distortion tolerances of 3-D objects is important. It is difficult to obtain mathematical formulas about the recognition tolerances. Instead, we used both digital simulation and optical experiments to evaluate the variation ranges of distortion such as scale, rotation and azimuth of objects. The experimental results are given in next section.

Recognizing a 3-D target with distortions needs a large number of sub-SDF filters for good correlation outputs of the OPR system. It is possible to storage many sub-SDF filters in the system memory. However, it is difficult to update a large number of filters in the LCSLM in real-time. Thus, we will focus on the 3D rotation distortions with small changes in scales or azimuths. This is true for images such as airplanes taken from a long distance. The distortion specifications for the OPR system are summarized as follows:

(1) shift-invariant: overall viewing field;
(2) rotation-invariant: $0^0 \sim 360^0$;
(3) azimuth-invariant: $36^0 \pm 2^0$;
(4) scale-invariant: $1 \sim 0.9$.

We found in our experiments that the number of SDF Filters is about 40 to realize the distortion-invariant recognition of targets. This optimizes and reduces the number of RFLs in the given scene.

2.3 Post target recognition approach using neural networks

In the post processing of correlation output of OPR system, the correlation decision making is based on the intensity of correlated peak and the suitable threshold. The use of thresholding can only eliminate low level noises in the noisy scenes. It can't reject high level noise sources or false objects while identifying all true auto-correlations correctly.

Artificial neural networks[5] process information in a parallel fashion and distributed mode. They operate in a robust, fault tolerant manner and can respond to environment changes through the use of well defined learning rules. However, for images with large number of pixels, the large computational loads required with ANNs make it difficult to perform recognition in real-time. Fortunately, in some applications, targets are of small sizes. After identifying the region of target, these targets can be recognized by ANNs in real-time.

To further improve the recognition reliability of real-time OPR, we proposed the two-step target recognition approach using artificial neural networks. Firstly, we use the mapping relation between the input plane (targets in scene) and output plane(correlation peak positions) in optical correlation filtering system to search the positions of correlation peaks above a given threshold; this will identify the region of interest (ROI) with small size window including the targets in the input plane. Secondly, artificial neural networks are used to further evaluate the targets in each ROIs. This process is reliable and could be implemented in real time, as each small ROI includes useful information about targets.

A three layer back-propagation neural network (BPNN) algorithm is designed for image recognition in scenes directly. We train the BPNN and adjust weights though off-line digital iterative processing using a supervised learning rule and a set of distorted image samples as training sets. The trained BPNN is then used to perform on-line target recognition tasks as follows:

(1) In the SDF based OPR system, the correlation peaks above a given threshold are identified and their locations are placed in output files.

(2) Based on mapping data tables of position between the input plane and output plane, we extract the ROI centered in the peaks in output files within a N×N rectangular window, where the size N is defined by experiments.

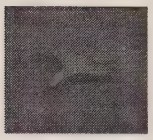
(3) Target images of ROIs are recognized by BPNNs in the personal computer.

3. EXPERIMENTAL RESULTS

3.1 The range of distortion-invariant image recognition

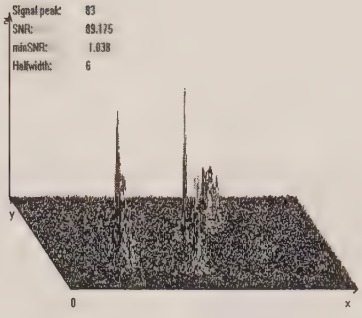
Our OPR system is tested in numerous scenes. A typical images containing 3D distorted objects such as fight plane F18, civil airplane A-320 and car and the corresponding results are shown in Figures 2(a) to 2(d). The car is used as a false target. Figure 2(a) is the original scene including several targets, 2(b) is the recognized target, 2(c) is the scene preprocessed by a digital edge operator, and 2(d) is the correlation outputs. The experimental results show the high SNR of correlation outputs.

(a)



(b)

(c)



(d)

Fig.2 Correlation recognition results in our OPR system

We further studied the effects of different distortion parameters of 3-D object on recognition in our OPR system. We used the fight plane F18 model and the civil airplane A320 model as examples to optimize sampling intervals of 3-D distortion object for construction of the SDF filters. The tolerance degree is decided by given enough high SNR of correlation outputs. Table 1 lists the experimental results about the tolerance degree of scale-invariant recognition for F18 model based on single match filter.

Table 1 Tolerance versus image scales

Size / Scale	32×32 pixels		64×64 pixels		128×128 pixels	
(scale rate)	Ave SNR	Min SNR	Ave SNR	Min SNR	Ave SNR	Min SNR
0.8	3.859	1.680	3.171	1.759	2.876	1.143
0.9	4.820	2.133	5.720	2.988	4.833	2.740
1.0	5.837	2.674	5.749	4.342	5.827	3.024
1.1					2.432	1.378
1.2						
Tolerance	0.2		0.2		0.2	

Tables 2 and 3 list the experimental results for F18 model about the tolerance degrees of azimuth-invariant recognition and of rotation-invariant recognition on a single match filter, respectively.

Table 2 Tolerance versus image azimuth angle

Azimuth range	0^0~30^0		35^0~60^0		
Azimuth angle	Ave SNR	Min SNR	Azimuth angle	Ave SNR	Min SNR
0^0	159.86	6.27	35^0	135.61	13.27
5^0	149.86	5.81	40^0	122.03	8.19
10^0	121.71	4.93	45^0	84.60	7.00
15^0	118.39	4.79	50^0	37.88	1.95
20^0	116.46	3.89	55^0	20.88	1.58
25^0	109.01	4.36	60^0	18.98	1.46
30^0	108.85	3.54			
Tolerance	30^0			25^0	

60^0~75^0			75^0~84^0			84^0~90^0		
Azimuth angle	Ave SNR	Min SNR	Azimuth angle	Ave SNR	Min SNR	Azimuth angle	Ave SNR	Min SNR
60^0	135.61	13.27	75^0	79.38	8.10	84^0	87.17	7.15
63^0	122.03	8.19	78^0	34.82	3.27	87^0	26.84	2.33
66^0	84.60	7.00	81^0	29.10	1.27	90^0	26.49	1.87
69^0	37.88	1.95	84^0	13.74	1.27			
72^0	20.88	1.58						
75^0	18.98	1.46						
tolerance	15^0			10^0			5^0	

Table 3 The tolerance versus distortion rotation angle

Image size / Rotation	32×32		64×64		128×128	
angle	Ave SNR	Min SNR	Ave SNR	Min SNR	Ave SNR	Min SNR
0^0	7.372	2.886	6.616	3.528	6.409	3.528
2^0	7.631	3.299	5.796	3.153	3.759	2.173
4^0	7.101	2.920	3.413	1.766	2.778	1.232
6^0	5.559	2.418	2.637	1.478	2.208	1.188
8^0	4.014	1.699	2.303	1.257		
10^0	3.469	1.357				
tolerance	10^0		8^0		6^0	

From the experimental results, we concluded that: (1)for scale-invariant recognition, the tolerance degrees are about 1:0.8, which indicates that a new match filter is needed to transform the distorted object (0.8 degree distortion) back to the original (without distortion); (2) for azimuth-invariant recognition, the tolerance degree will decrease as the azimuth angle of object increases; (3) for rotation-invariant recognition, the tolerance degree is different for images of different sizes; the rotation sampling interval angle is about 10 degree for 32×32 pixels image, 8 degree for 64×64 pixels image, and 128×128 pixels image. Therefore, a rotation-invariant SDF filter is generated with 4 group training targets, while each group includes 9 images with 32×32 pixels and 10 degree rotation angle intervals.

As for the rotation distortion of an object, typical SDFs generated from the three different groups of training images (F18 model) are given in figure 3. Our experiments show that only eight training images with rotation distortion are needed to construct SDF filters with good correlation SNR. In practice, we can construct a few of SDF filters to perform the distortion-invariant recognition by renewal SLM in real-time, using constraint range of distortion based on the recognized task.

(a) (b) (c)

Fig.3, SDF generated by :(a) 4 rotation images, (b) 8 rotation images ,(c) 16 rotation images.

3.2 Second time recognition results using ANN

From the experimental results, we found that some targets are misclassified by using single threshold method under the strong noise or false target. Therefore, we employ the BP neural networks approach to perform the 3-D target recognition again. The details are discussed in section 2.2. Note that same distortion training images with 32×32 pixels are used for both SDF filters and adjustment weights of ANN. The comparison of experimental results with ANN and without ANN in the OPR system is shown in Table 4. As can be seen in Table 4, the correct rate of recognition for 3-D distortion images increased with both threshold technique and artificial neural networks, although a little extra computation and slightly reduced recognition rates are also observed. These are acceptable for optical implementation in real-time. It is reported[6] that the hardware of ANN based on the ZISC technology invented by Guy Pailet has been commercialized, which will improve the ability further to process data and target recognition if used in our OPR system.

Table 4 The experimental results for airplane A-320 model

Results / Method	Measuring images (times)	Wrong recognition (times)	Correct rate of recognition (%)	Renewal rate (frame/s)	Recognizing rate (times/s)
Simple threshold	800	108	87.5.0	25.0	6.9
Threshold & ANN	800	16	98.0	24.5	2.3

4. CONCLUSION

In this paper, we studied the tolerance degrees of rotation-invariant recognition. We proposed a new OPR system for distorted target recognition. In the proposed OPR system, we optimized the reference filter Libs and SDF filters and reduced the number of SDF filters for 3-D distortion-invariant recognition using constraint range of distortion based on the practical recognized tasks.

We also applied the mapping relation to the input (object region) and the output(correlation peak). The region of interest (ROI) with small size window including the target was processed and a post recognition stage was performed by artificial neural networks to further improve the reliability of our proposed system. The Real-time automatic optical target recognition was realized by temporal multiplexing technique in renewal SDF filters, which performed with electronic addressing spatial light modulator. The experiment results show the efficieny of the proposed OPR system.

REFERENCES

(1) D. Casasent, W. Rozzi, "Projection Synthetic Discriminant Function Performance", Optical Engineering, 23, 716-720 (1984)

(2) Joseph J. Booth, U.S. Army Missile Command, "Automatic post processing of correlation plane imagery", Proc. SPIE, Vol. 2490, 108-116 (1995)

(3) A. Mahalanobis, B.V.K. Vijaya Kumar, and D. Casasent, "Minimum average correlation energy fileters", Appl. Opt., Vol. 26 ,3633-3640 (1987)

(4) A. Mahalanobis, B.V.K. Vijaya Kumar, S. Song, S.R.F. Sims, and J.F. Epperson, "Unconstrained correlation filters", Appl. Opt., Vol. 33, 4465-4471 (1994)

(5) J.Hertz, A.Krogh, R.Palmer, Introduction to the Theory of Neural Computation, Addison-Wesley Publishing Company, 1991

(6) ZISC Technology Reference Guide (version), Silicon Recognition, Inc.

Incoherent acoustooptic image correlator with the kinoform

Sergey N. Starikov, Vladislav G. Rodin, Ivan V. Solyakin, Ekaterina A. Shapkarina, Alexander P. Chervonkin

Moscow Engineering Physics Institute (State University), Russia, Moscow

ABSTRACT

Fourier holograms are commonly used for reference images storing in diffraction correlators with spatially coherent or spatially incoherent illumination. Kinoforms can be a real alternative to Fourier holograms in the correlators. The kinoform represents a computer-synthesized optical element which performs only a phase modulation of a light wave. The kinoform restores true intensity of the recorded image and random distribution of phase. Therefore, it can be utilized for storing reference images, first of all, in correlators with spatially incoherent illumination. The absence of carrier frequency reduces demanded number of pixels of the spatial light modulator being used. Since the kinoform provides reconstruction of reference image in zero diffraction order, requirement on monochromaticity of illumination are decreased as well. The diffraction correlator with the kinoform used as spatial frequency filter is considered. The 2-D acoustooptic deflector was employed to form input images in real time by monochromatic spatially incoherent light. The reference images were recorded on the commercially available kinoforms. The input and reference images were of 256×256 pixels and 200×200 pixels respectively. Since input images were consisted of approximately 400 pixels with non-zero brightness, the image update frequency was gained at 200 Hz. The experimental setup and experimental results on images recognition are presented.

Keywords: Incoherent correlator, image recognition, kinoform, acoustooptic deflector

1. INTRODUCTION

In diffraction correlators with monochromatic spatially coherent[1,2] or spatially incoherent[3-5] illumination Fourier holograms as spatial frequency filters are generally employed. Another approach is to use kinoforms[6] instead of Fourier holograms for reference images storing in diffraction correlators. The kinoform is a computer-synthesized optical element, performing only a phase modulation of a light wave. Applications of kinoforms include laser scanning and machining, interferometry, optical interconnections, and synthesis of novel elements for optical information processing. The kinoform usually restores true intensity distributions of the image recorded with pseudo-random phase distributions. Therefore, it can be utilized for storing reference images, first of all, in correlators with spatially incoherent illumination. The diffraction efficiency of the kinoform is close to 100 %. The absence of carrying spatial frequency subsequently reduces requirements to the number of pixels of spatial light modulator used for reference images recording. It allows to create operatively controlled filters for image recognition using the commercially available element base[7]. Moreover, because of image restores from kinoform in zero order the requirements for monochromaticity of illumination are reduced.

Several studies on employing kinoforms as spatial frequency filters in coherent correlators have been reported[8-10]. There are following problems when spatially coherent illumination is used. If reference image is recorded on the kinoform under control only intensity distribution, then the reconstructed reference image has a pseudo-random phase distribution. Therefore, no proper correlation with input images could be obtained in spatially coherent light. Otherwise, to synthesize the kinoform under control both intensity and phase distributions of the reference image, notable part of pixels must be granted for a dummy area around the reference image. The dummy area contains pattern with pseudo-random intensity and phase distributions and provides the equalization of amplitudes of total Fourier spectrum. As a result, a spatial light modulator with great number of pixels is demanded and the useful diffraction efficiency of the kinoform, corresponding to the reference image, is small.

Optical Pattern Recognition XV, edited by David P. Casasent, Tien-Hsin Chao, Proc. of SPIE Vol. 5437 (SPIE, Bellingham, WA, 2004) · 0277-786X/04/$15 · doi: 10.1117/12.541780

Thus, as it was noted above, advantages of kinoforms are realized completely in correlators with spatially incoherent illumination, when the correlation of images intensities (not complex amplitudes) is calculated. Moreover, the utilization of spatially incoherent light reduces requirements on positioning of optical element and decreases influence of phase noises.

The use of two-coordinate acoustooptic deflector (2-D AOD) is one of ways to achieve efficient information input to the diffraction correlator with spatially incoherent monochromatic illumination [11,12]. The principle of work of the correlator consists in the following. In each moment of a time AOD forms a plane wave, corresponding to one point of the generated input image. This wave reconstructs the reference image, recorded on Fourier hologram or kinoform, in the plane of photosensor array. The correlation signal is formed as sum of intensities of shifted and weighted reference images, reconstructed by all points of input image.

2. EXPERIMENTAL SETUP OF IMAGE CORRELATOR WITH 2-D AOD AND KINOFORM

2.1. Scheme of diffraction image correlator with monochromatic spatially incoherent illumination

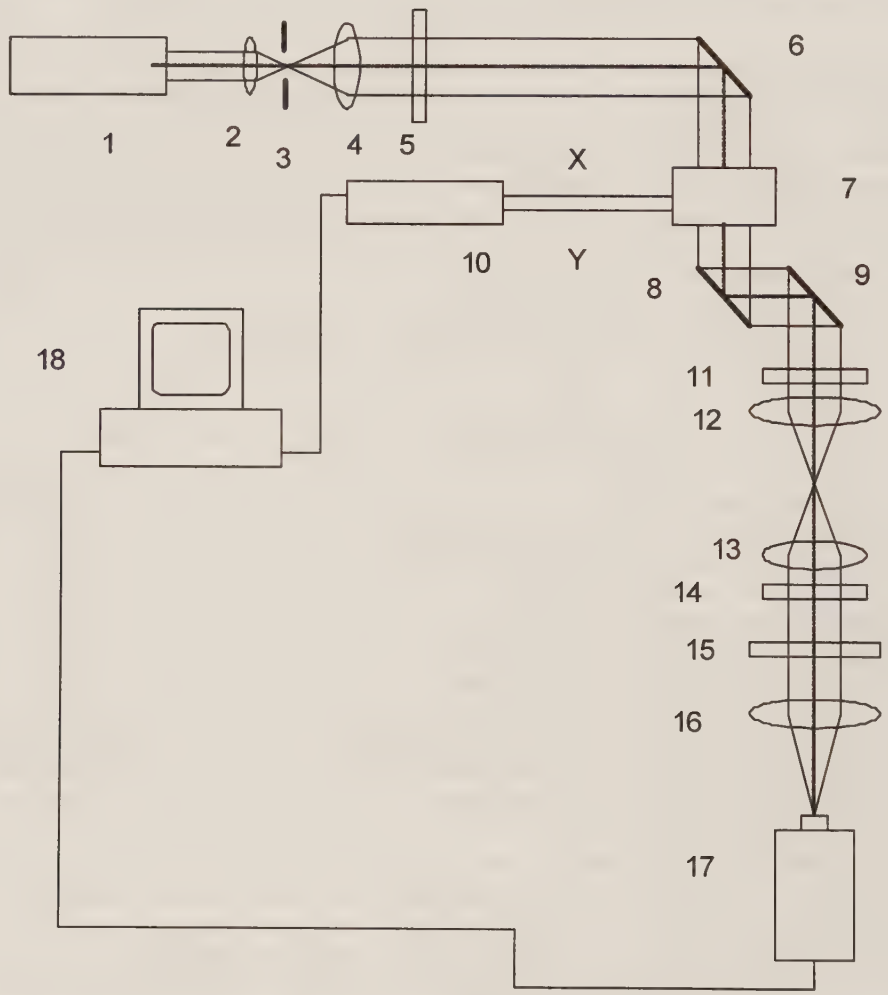


Fig. 1. The experimental setup of image correlator with 2-D AOD as input device and kinoform as spatial frequency filter: 1 - laser, 2 - microobjective, 3 - diaphragm, 4 - objective, 5 - polarizer, 6 - mirror, 7 - 2-D AOD, 8 and 9 - mirrors, 10 - generator of 2-D AOD, 11 - polarizer, 12 - objective, 13 - objective, 14 - polarizer, 15 - kinoform, 16 - objective, 17 - video camera, 18 – computer

Schematic view of correlator setup is shown in Fig. 1. Light from He-Ne laser 1 passes through the beam expander (microobjective 2, diaphragm 3 and objective 4), polarizer 5 and adjusted mirror 6. Expanded beam illuminates the two-coordinate acoustooptic deflector 7. The computer 18 controls 2-D AOD through the two-channel generator 10. The light, forming input image, is directed by mirrors 8 and 9 in optical system of the correlator. The polarizer 11 serves for background light elimination. Spatially incoherent input image generated by 2-D AOD is focused in back focal plane of the objective 12. This plane is also input plane of the correlator and front focal plane of the objective 13. The telescope consisting of objectives 12 and 13 is used for scale matching of input images and reference image, recorded on kinoform 15. Objective 16 focuses shifted and weighted reference images, reconstructed by all points of input image on CCD video camera 17. The polarizer 14 serves for varying intensity falling on CCD. The correlation signal is formed as a result of time integration in CCD camera. Correlation signals with 512×512 pixels are captured by computer 18 for further processing.

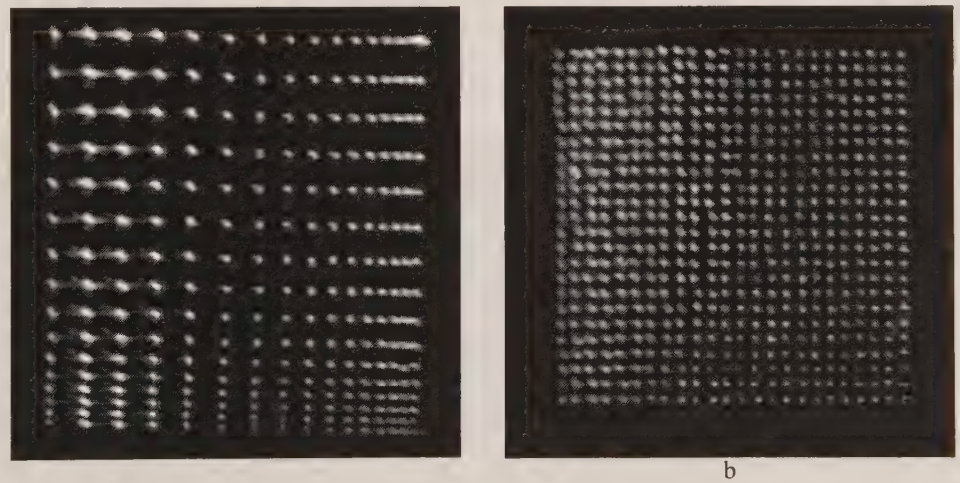
2.2. Two-coordinate acoustooptic deflector (2-D AOD) as input device for image correlator

In our experiments TeO_2 two-coordinate acoustooptic deflector, developed at North-Osetian State University (Russia) was used [11]. The main technical characteristics of the 2-D AOD are given in Table 1.

Table 1. Characteristics of TeO_2 two-coordinate acoustooptic deflector

1.	Frequency bandwidth	50 MHz
2.	Duration of non-zero brightness pixel formation	< 10 μs
3.	Number of pixels in image	256×256
4.	Diffraction efficiency	40%
5.	Input aperture	6 mm

Distortion of raster's pitch of AOD was equal approximately 20% (see Fig. 2a). Programmable correction of distortion is applied while input image for AOD is generated in computer. Programmable correction reduces distortion of raster's pitch to 5% (see Fig. 2b).



a b

Fig. 2. The raster formed by the acoustooptical deflector without (a) and with (b) programmable correction of distortion of raster's pitch

The input images in experiments were of 256×256 pixels in size with at about 400 pixels of non-zero brightness. Thus, the input image update frequency was gained at 200 Hz.

2.3. Kinoforms with recorded reference images

Two commercially available static kinoforms were used in experiments. Reference images reconstructed from kinoforms are shown in Fig. 3 (reference object "1") and Fig. 4 (reference object "2"). Each kinoform consists of periodically repeated identical subkinoforms. Such replication is necessary to reduce speckle-effect in reconstructed image from kinoform. The main characteristics of kinoforms were measured and are presented in Table 2.

Fig. 3. The reference object "1" reconstructed from kinoform

Fig. 4. The of reference object "2" reconstructed from kinoform

Table 2. Characteristics of kinoforms

1.	Kinoform size	2.7×2.7 mm^2
2.	Replicated subkinoform size	0.9×0.9 mm^2
3.	Size of resolution element	4.5×4.5 μm^2
4.	Diffraction efficiency	> 50%
5.	St. deviation of '1' pixels in reconstructed images	0.23
6.	St. deviation of '0' pixels in reconstructed images	0.03

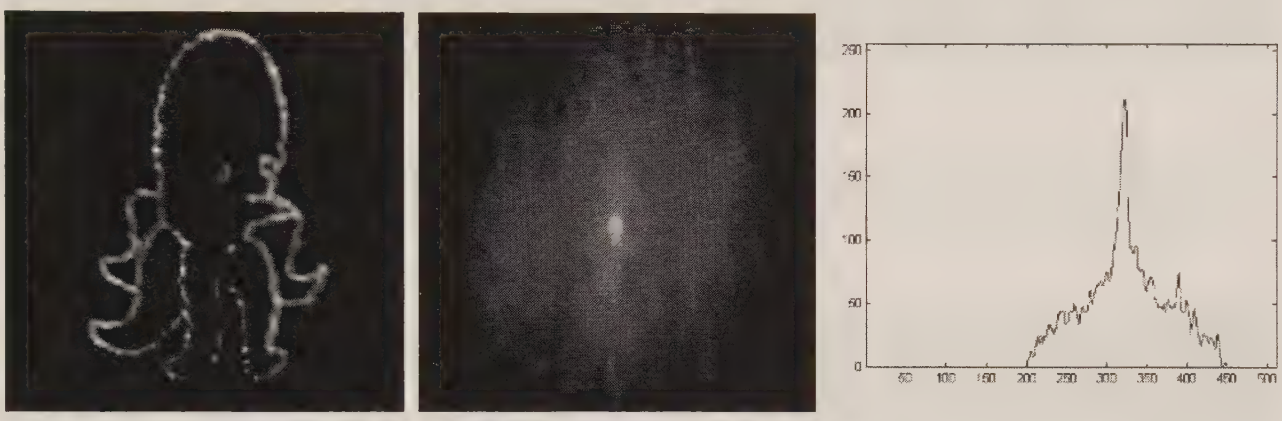
Reference binary images recorded on kinoforms had total number of pixels 600×600, but actual one was of 200×200 because of the replication.

3. EXPERIMENTAL RESULTS

Experiments on images recognition were performed on setup of correlator with 2-D AOD as input device and kinoform as spatial frequency filter. Input images of test objects "1" and "2" similar to reference objects "1" and "2" (see Fig. 3 and Fig. 4) were generated by computer, which controls 2-D AOD. Images of test objects "1", "2" and the sum of them were formed by 2-D AOD in front focal plane of the objective 13 (see Fig. 1). Figures 5 and 6 show input images (a), output signals of the correlator (b) and profiles of central cross-section of correlation signals (c) when the kinoform with reference object "1" was used. Figures 7 and 8 illustrate correlation signals in the case when kinoform with reference object "2" was placed in the correlator. Experimental values of peak and sidelobe values of correlation signals and comparison of experimental and computer simulated values of peak-to-sidelobe ratio are given in Table 3.

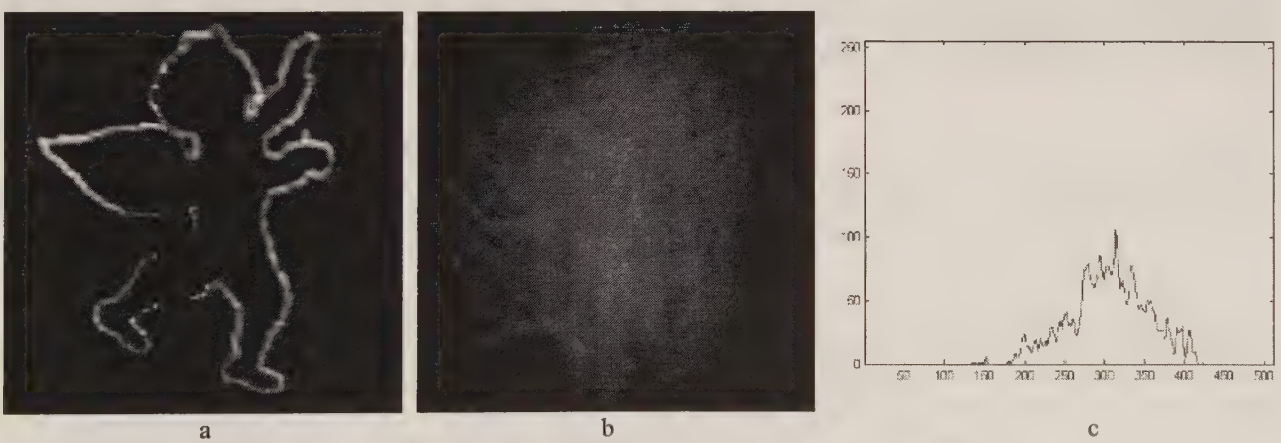
Table 3. Peak and sidelobe values of correlation signals

Input object	Reference object	Peak (experimental)	Sidelobe (experimental)	Peak/sidelobe (experimental)	Peak/sidelobe (calculated)
"1"	"1"	212	99	2.14	2.83
"2"	"1"	103	100	1.03	1.02
"2"	"2"	248	89	2.79	3.07
"1"	"2"	103	97	1.06	1.02



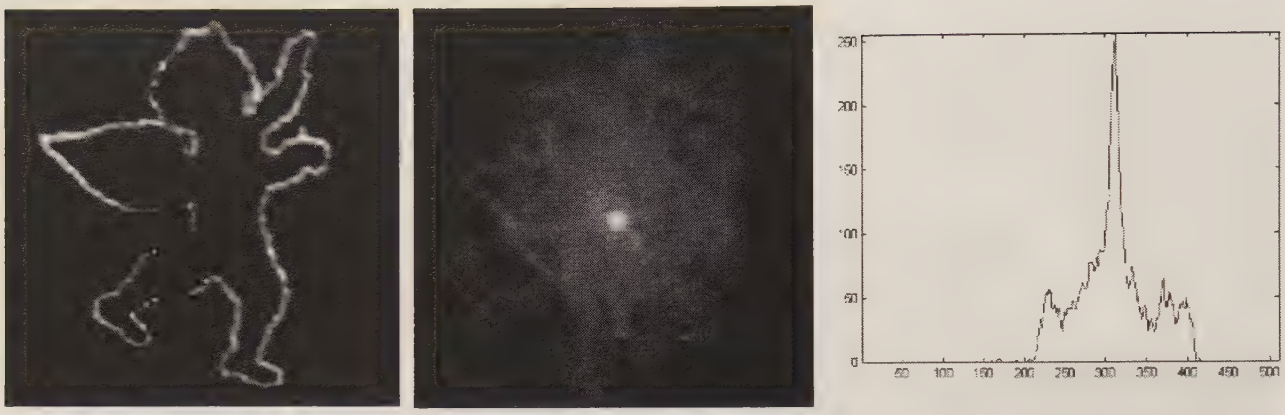
a b c

Fig. 5. Experimental results on image recognition: a – the input image of test object "1" formed by 2-D AOD; b - the correlation signal of test object "1" and reference object "1" recorded on kinoform; c - the profile of central cross-section of correlation signal



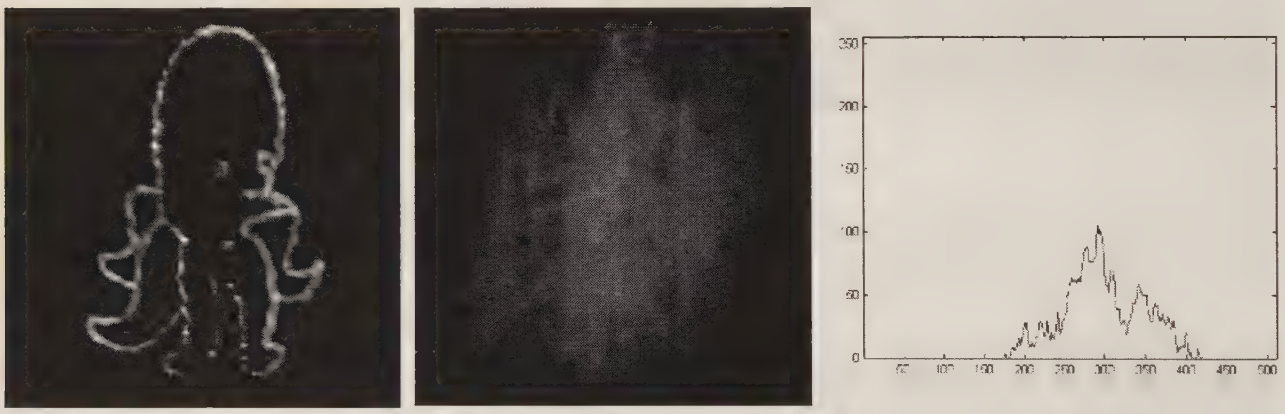
a b c

Fig. 6. Experimental results on image recognition: a – the input image of test object "2" formed by 2-D AOD; b - the correlation signal of test object "2" and reference object "1" recorded on kinoform; c - the profile of central cross-section of correlation signal



a b c

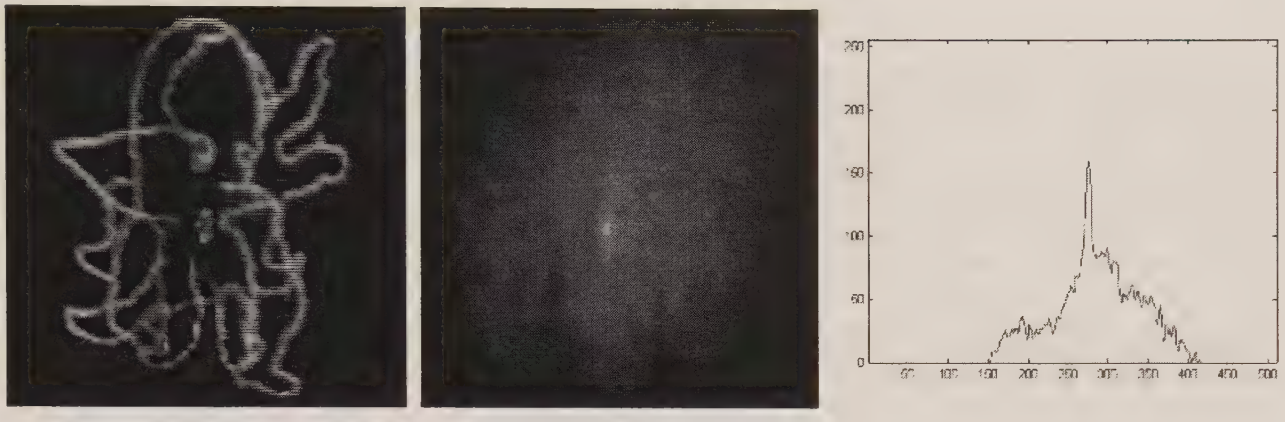
Fig. 7. Experimental results on image recognition: a – the input image of test object "2" formed by 2-D AOD; b - the correlation signal of test object "2" and reference object "2" recorded on kinoform; c - the profile of central cross-section of correlation signal



a b c

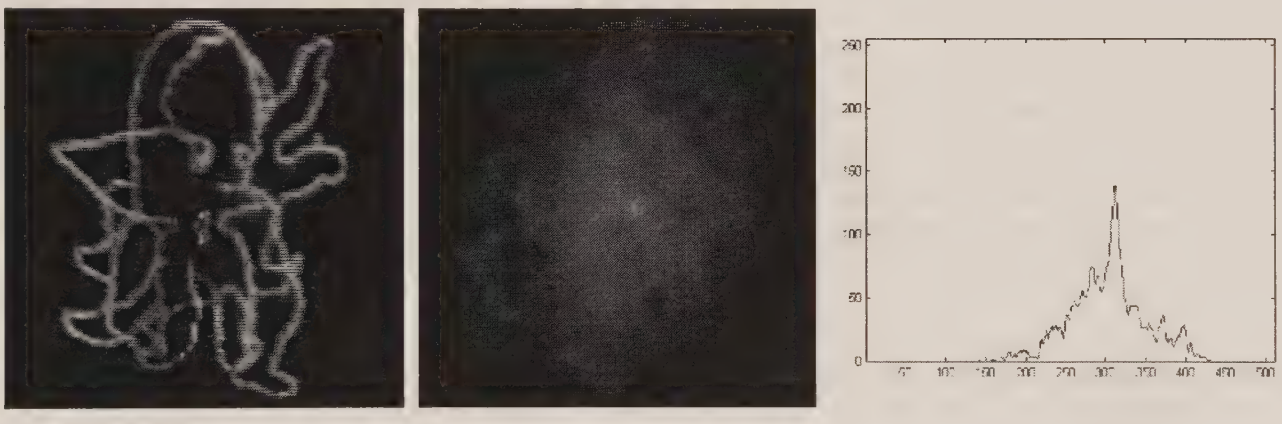
Fig. 8. Experimental results on image recognition: a – the input image of test object "2" formed by 2-D AOD; b - the correlation signal of test object "1" and reference object "2" recorded on kinoform; c - the profile of central cross-section of correlation signal

In some experiments image of the overlapping of both test objects "1" and "2" were inputted in the correlator simultaneously. Test object "1" was recognized when the kinoform with reference object "1" was used (see Fig. 9). Value of peak-to-sidelobe ratio was equal to 1.58. And test object "2" was recognized too when the kinoform with reference object "2" was used (see Fig. 10). Value of peak-to-sidelobe ratio was equal to 1.62 in this case.



a b c

Fig. 9. Experimental results on image recognition: a – the input image is the overlapping sum of test object "1" and"2"; b - the correlation signal of the sum of test objects and reference object "1" recorded on kinoform; c - the profile of central cross-section of correlation signal



a b c

Fig. 10. Experimental results on image recognition: a – the input image is the overlapping sum of test object "1" and"2"; b - the correlation signal of the sum of test objects and reference object "2" recorded on kinoform; c - the profile of central cross-section of correlation signal

The experimental results demonstrate the ability of the investigated correlator to calculate correlation functions of images formed by the acoustooptic deflector and images recorded on the kinoform. Exceeding of signals of coincidence above signals of difference was not less than twice for each test objects and not less one and a half for the overlapping sum of test objects. The values of signals are agreed with results of computer simulation.

4. CONCLUSIONS

Described diffraction correlator with reference image stored on a kinoform allows to perform real time processing of input images formed by 2-D AOD. The input images were of 256×256 pixels in size with at about 400 pixels of non-zero brightness, so the image updating (correlation computing) frequency gained at 200 Hz. Reference images recorded on kinoforms had total number of pixels 600×600, but actual one was of 200×200 approximately. Experimental results on recognition of test objects are in good agreement with those ones of computer simulation. The use of kinoform instead Fourier hologram provides high diffraction efficiency, possibility of employment of spatial light modulators with restricted number of pixels for operatively controlled filters, and to reduce requirements for monochromaticity of illumination. Since input illumination is spatially incoherent, then the correlator is less sensitive to phase aberrations and to displacements of optical elements. The using of semiconductor laser as the light source and removing of two Fourier transform cascades approve to decrease the correlator dimensions significantly.

ACKNOWLEDGMENTS

This work was supported by the Ministry of Education of the Russian Federation (grant no. A03-2.9-839 for support of research work of post-graduate students).

REFERENCES

1. A. Vander Lugt, "Signal detection by complex filtering," *IEEE Trans. Inform. Theory* IT-10, pp. 139-145, 1964.
2. H. Sjoberg, B. Noharet, L. Wosinski, R. Hey, "Compact optical correlator: preprocessing and filter encoding strategies applied to images with varying illumination," *Opt. Eng.* 37(4), pp. 1316-1324, 1998.
3. A. W. Lohmann, "Matched filtering with self-luminous objects," *Appl. Opt.* 7(3), pp. 561-563, 1968.
4. O. I. Potaturkin, V. I. Khotzkin, "Diffraction correlators invariant to the shift of optical elements," *Opt. Commun.* 34(2), pp. 159-163, 1980.
5. Yu. A. Bykovsky, A. A. Markilov, V. G. Rodin, S. N. Starikov, "Optical information processing with transformation of the spatial coherence of light," *Quantum Electronics* 25(10), pp. 1014-1019, 1995.
6. L. B. Lesem, P. M. Hirsch, and J. A. Jordan, Jr., "The kinoform: a new wavefront reconstruction device," *IBM J. Res. Dev.* 13, pp. 150-155, 1969.
7. J. Amaco and T. Sonehara, "Kinoform using an electrically controlled birefringent liquid-crystal spatial light modulator," *Appl. Opt.* 30(32), pp. 4622-4628, 1991.
8. G. Paul-Hus and Y. Sheng, "Optical real-time kinoform for on-axis phase-only correlation using a liquid-crystal television," *Proc. SPIE* 2043, pp. 287-295, 1994.
9. I. Moreno, J. Campos, C. Gorecki, and M. J. Yzuel, "Effects of amplitude and phase mismatching errors in the generation of a kinoform for pattern recognition," *Jpn. J. Appl. Phys.* 34, pp. 6423-6432, 1995.
10. M. Starzomski and M. Cada, "Application of a liquid-crystal display as a multi-level quantized phase-mostly filter in a 4-f optical correlator," *Proc. SPIE* 3491, pp. 1135-1140, 1998.
11. S. A. Amelyanov, Yu. A. Bykovsky, E. N. Eloev, A. A. Markilov, S. N. Starikov, M. V. Terekhov, "Incoherent holographic image correlator with 2-D acoustooptic deflector as input device," *Proc. SPIE* 3900, pp. 334-337, 1999.
12. Sz. Takes, L. Orzo, T. Roska, "Design aspects of an optical correlator based CNN implementation," *Proc. IEEE International Symposium on Circuits and Systems*, Vol. 1, pp. 353-356, 2001.

DESIGN OF THE OPTICAL-ELECTRONIC COMPUTER SYSTEMS FOR IMAGE PROCESSING IN STATIONARY AND DYNAMIC MODES

Veacheslav L. Perju[1], David P. Casasent[2]

[1]Technical University of Moldova, Stefan Mare av., 168, Chisinau, MD-2012, Republic of Moldova
Tel: (373)79431245, E-mail: perju@adm.utm.md

[2]Carnegie Mellon University, Department of Electrical and Computer Engineering, Pittsburgh, PA 15213 USA, Tel: 412-268-2464, E-mail: casasent@ece.cmu.edu

ABSTRACT

The theory of designing the optical-electronic image processing computer systems has been presented. A model of parallel image processing system has been considered, that is based on the principle of function decomposition. A structure model of computer system has been examined, that is a conveyor of parallel processors. The implementation possibilities of different image processing operations by optical and electronic computer means have been analyzed. Evaluation of time outlay in the system, while processing an image or a series of them has been made. There have been exposed the dependence of image processing time from conveyor length change and the correlation of optical and electronic devices. The designing method of image processing systems in static mode has been described. The results of investigations of the influence of the median square deviation, of the processing time in the modules on the throughput capacity of the system under the different electronic and optical modules quantity are presented. The recommendations of increasing the system's throughput capacity are formulated. On the bases to these recommendations, the system design method of image processing in the dynamic mode is elaborated.

Keywords: computer system, conveyor, electronic, image, optical, processor

1. INTRODUCTION

Over the last years a significant progress was achieved in the development of special and general - purpose electronic multiprocessor computer systems[1-3]. Such systems are characterized by a high productivity (up to hundreds billions operations per second). Although they are complicated and consume a significant power that in some cases limits the employment sphere.

The significant results have been obtained in the construction of the optical and optical-electronic processors with acceptable weight - size characteristics and high through - put[4,5]. High productive processors, which implement different operations of preliminary image processing, image features extraction and invariant pattern recognition have been developed.

However the analysis shows that not all of the image processing operations can be implemented at present by the help of the optical processors. Besides, optical computers are characterized by the limited flexibility. Therefore the construction of high productive efficient computer systems for image processing is connected with the development of multiprocessor systems, which are based on the combination of optical and electronic structures, optimal distribution of the solving problem function among the processing devices of different type.

A number of questions, which are connected with investigation and development of the theory of image processing optical-electronic computer systems design are being presented in this paper.

In section 2 a model of parallel image processing system have been considered. In section 3 the generalized structure of parallel image processing computer system and the implementation possibilities of different image processing operations with the help of optical, electronic computer means are presented. In section 4 there have been made time expenditures in the system and the through-put at the processing of one and series of images. In section 5 there have been presented the method of image processing system design in static mode at different processing devices resources. In section 6 are presented the results of investigations of the influence of the median square deviation, the influence of time of the

Optical Pattern Recognition XV, edited by David P. Casasent, Tien-Hsin Chao, Proc. of SPIE Vol. 5437 (SPIE, Bellingham, WA, 2004) · 0277-786X/04/$15 · doi: 10.1117/12.548652

processing in the modules on the throughput capacity of the system under the different electronic and optical modules quantity. According to the results of investigations the recommendations of increasing the system's throughput capacity are formulated. On the bases to these recommendations, the system design method of image processing in the dynamic mode is elaborated. An example of image processing computer system design is presented in section 7.

2. MODEL OF PARALLEL IMAGE PROCESSING

Let the image processing problem is determined in the following way: $W\{P(x,y)\}\rightarrow RO$, where W - is the function of image processing; P(x,y) - initial image; RO - the processing result, which can be an image or numerical data.

The most of the image processing problems can be represented in the form of a linear sequence of separate sub problems. The decomposition principle utilization allows to organize parallel processing of two types: in time and in space. Parallel processing in time consists in organization of the sub problems conveyor, which are implemented consecutively.

The spatial concurrency means that on each step of the conveyor processing, the parallel processing of the whole image or its separate fragments is organized. The last type of concurrency is defined by the fact that we can point out some classes of operations while proceeding the image processing. They are executing at each of the image pixels, over separate regions of the image or over the whole image. In other words, to obtain the result we need local (as in the first two cases) or global (as in the last case) initial information. At the working with the local information, the image can be represented as a set of sub images, which are processed by the corresponding set of processor units. The spatial concurrency can be implemented by the help of processor units set of electronic or optical type.

Corresponding with the above-mentioned, the model of the parallel processing of the images can be represented in the following way:

$$RO = C_h\{\bigcup_i [W^j_i(P_{ij})]\},$$

where C_h - the superposition operation, i - the step number of the conveyor; U - the processing results union operation during one of the conveyor steps; j - the image fragment number; Wji- the operation of the processing of the j-th image fragment on i-th step; Pij- the j-th image fragment on i-th step.

This model can be also represented in the form of the operator:

$$P(x,y)=>\{W1\}_{M1}\rightarrow\{W2\}_{M2}\rightarrow...\rightarrow\{Wk\}_{Mk}=>RO.$$

The peculiarity of the presented model consists in the maximum level of disparalleling of image processing operations that ensure the high productivity of the computer system, which implements this model. The productivity also depends on element base which is utilized. In connection to this fact let us look through the generalized structure of parallel image processing computer system and the implementation possibilities of different image processing operations applying optical and electronic computer means.

3. GENERALIZED STRUCTURE OF PARALLEL IMAGE PROCESSING COMPUTER SYSTEM

According to the proposed model, the generalized structure of parallel image processing computer system can be represented as shown in Fig.1, where Yi - is a device that implements the operation W_i, B_i - buffer device. The device Y_i may be optical or electronic. In the last case it can contain one or a group of processor elements that function in parallel way.

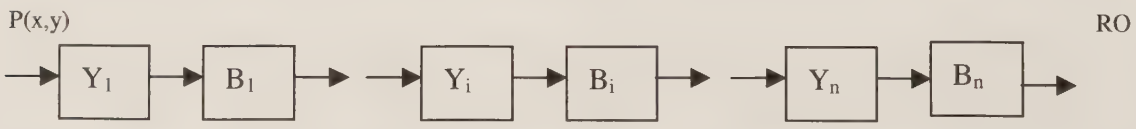


Fig.1. The structure of the image processing computer system

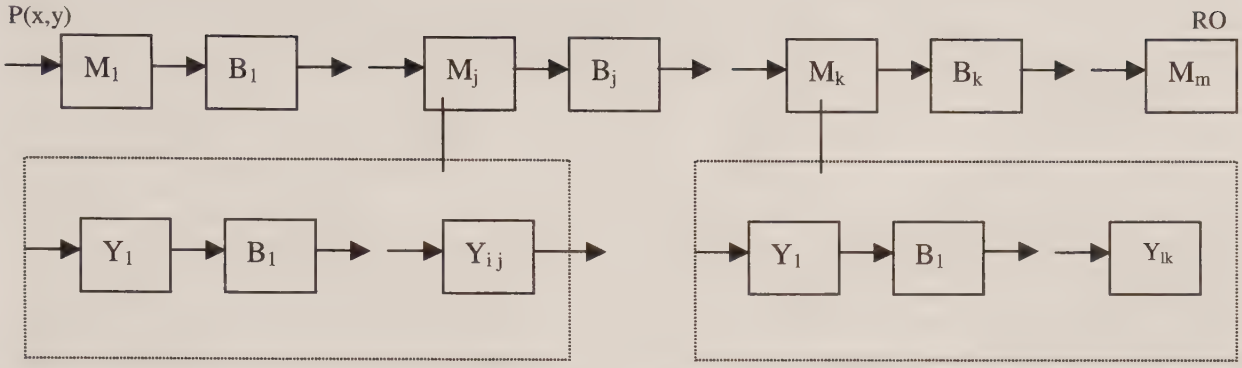


Fig.2. The module structure of the image processing computer system

The utilization of optical or electronic processing devices during different steps of the image processing conveyor system will be defined by some factors.

At first, from the way of initial image forming. The image can be formed by the help of electronic or optical means. The analysis of image forming units (IFU) characteristics showed that electronic and optical IFU provide the identical resolution ability. The difference between them is in implement forming speed. Optical IFU allow to form images 2-8 times faster.

The next factor is the implementation possibility of the operations by the help of the optical means. In the Table 1 the possibilities and the execution time of the image processing operations by the help of the different computer means are presented.

The analysis of the data, presented in the Table 1 shows that the time of optical implementation of some image processing operations is much less than one of their implementation by electronic means.

While using an optical processor it is expedient to give the full optical image to the input then on exit we shall have also a formed full optical image. In the case of implementation of an operation in an electronic processing device we can give the information about full image or about its fragment to the input, i.e. in the first case the information is put in/out in parallel way, optically, but in the second case it is done either in parallel way or consecutively in the form of electronic signals.

The structures of communication between two different steps are defined in the dependence on the processors types. Let us denote the buffer storage as $B_{m,n}$, where m,n are taking the values 1 or 2. Let us consider that number 1 corresponds to an electronic device, and number 2 - to an optical one. For example, the device $B_{1,2}$ is a buffer storage for connection of electronic and optical processors.

The time outlay during the signal buffering of different types is represented in Table 2.

The analysis of the possibilities of image processing operations implementation allowed to catch out the following:

1.The significant part of the wide used operations can be implemented optically.
2.The order of time outlay for optical implementation is 2-4 times smaller than for electronic one.
3. At present there exist some operations, which can't be implemented optically.

Table 1

The possibilities and the time of the implementation of the image processing operations

Image processing operations	Computing means		
	Optical	Optical-electronic	Electronic
1. Image forming	10ms	30ms	20(40)ms
2. Noises removal	10μs		172ms
3. Edge detection	10μs		172ms
4. Object extraction	5ms		2.41sec
5. Image transformation from Cartesian system of coordinates into polar or logarithmic - polar one	10μs	+	20(40)ms
6. Image rotation and scaling with arbitrary parameters	-	0.13sec	12sec
7. Calculation of two-dimensional Fourier transformation	10μs	+	0.5sec - 4min
8. Calculation of two-dimensional correlation and convolution functions	10μs	+	1sec - 8min
9. Calculation of geometrical moment features	10μs	+	+
10.Calculation of central moment features	-	0.6ms	0.2sec
11.Calculation of invariant moment features	-	2.8ms	0.25sec
12. Chords transformation of the image	10μs	+	+
13. Image features processing and comparing with a standards	-	+	2.7sec
14.Correlation fields analysis, calculation of coordinates maxima	-	10ms	20(40)ms
15.Computing process control	-	+	+

Table 2

Time outlay during the signal buffering of different types, ms

Kind of the buffer	Image resolution, pixels	
	512x512	256x256
$B_{1,1}$	40	20
$B_{1,2}$	50	30
$B_{2,1}$	40	20
$B_{2,2}$	10	10

In accordance to the above it is expedient to construct parallel optical-electronic image processing computer systems.

4. THE THROUGH-PUT OF THE IMAGE PROCESSING COMPUTER SYSTEM

Let us evaluate the time outlay in the system while processing one and series of images. In the structure represented in Fig.1 we can join the consecutively situated optical devices into optical processing modules and electronic devices - into electronic processing modules (Fig.2). Let the system consists of n_o optical and n_e electronic modules ($n_o+n_e=m$), and each module contains accordingly l^o_j optical and l^e_q electronic processing devices. The whole number of optical l^o, electronic l^e devices and also the conveyor length l_k will be:

$$l^o=\sum_{j=1}^{n_o} l^o_j, \quad l^e=\sum_{q=1}^{n_e} l^e_q, \quad l_k=l^o+l^e.$$

Let us consider the following time characteristics: t^o_{pj} - processing time in j-th optical module; t^e_{pq}- processing time in q-th electronic module; t^o_{kj} - switching time between devices in optical module; t^b_{oo} - buffering time of two neighbor optical processing devices; t^e_{kq}- switching time between devices in electronic module; t^b_{ee}- buffering time of two neighbor electronic processing devices; t^n_{kq}- switching time between the processors in electronic module devices; t^k_{oe} - time of buffering of optical modules with electronic ones; t^k_{eo} - time of buffering of electronic modules with optical ones. These parameters are estimated in the following way:

$$t^o_{pj}=\sum_{z=1}^{l^o_j} t^o_{pjz}, \quad t^e_{pq}=\sum_{f=1}^{l^e_q} t^e_{pqf}, \quad t^o_{kj}=(l^o_j-1)t^b_{oo}, \quad t^e_{kq}=(l^e_q-1)t^b_{ee}$$

$$t^n_{kq}=\sum_{f=1}^{l^e_q} t^n_{kqf},$$

$$t^k_{oe}=\sum_{i=1}^{n} t^b_{oei}, \tag{1}$$

$$t^k_{oe}=\sum_{i=1}^{n-1} t^b_{eoi}. \tag{2}$$

The value of n in expressions (1), (2) is defined in the following way:

$$n=\begin{cases} m/2 & \text{if } n_o = n_e \\ n_e & \text{if } n_o > n_e \\ n_o & \text{if } n_o < n_e. \end{cases}$$

The time outlay in optical and electronic modules can be described as following:

$$t^o_j = t^o_{pj} + t^o_{kj} = \sum_{z=1}^{l^o_j} t^o_{pjz} + (l^o_j-1)t^b_{oo},$$

$$t^e_q = t^e_{pq} + t^e_{kq} + t^n_{kq} - \sum_{f=1}^{l^e_q} t^e_{pqf} + (l^e_q-1)t^b_{ee} + \sum_{f=1}^{l^e_q} t^n_{kqf}.$$

The time outlay in optical-electronic computer system while processing one image can be defined as:

$$T = \sum_{j=1}^{n_o} t^o_j + \sum_{q=1}^{n_e} t^e_q + t^k_{oe} + t^k_{eo} =$$

$$= \sum_{j=1}^{n_o}\sum_{z=1}^{l^o_j} t^o_{pjz} + \sum_{q=1}^{n_e}\sum_{f=1}^{l^e_q} t^e_{pqf} + \sum_{j=1}^{n_o}(l^o_j-1)t^b_{oo} + \sum_{q=1}^{n_e}\{\sum_{f=1}^{l^e_q} t^n_{kqf} + (l^e_q-1)t^b_{ee}\} + \sum_{i=1}^{n}\{t^b_{oei} + t^b_{eoi}\}. \tag{3}$$

The analysis shows that the time outlay for image processing consists of two components - direct processing time and switching time. Switching time is defined by the time of data transfer (or intermediate image time) between conveyor steps (devices buffering time) and the time of processors switching in electronic processing devices.

Let t_{pi} is the time of image processing on i-th step of the conveyor, t_{ki} is the switching time. The whole processing time on the i-th step of the conveyor will be:

$t_i = t_{pi} + t'_{ki}$, where $t'_{ki} = \max\{t_{ki-1}, t_{ki}\}$.

The through-put of a conveyor while processing a series of images is defined as[1]:

$$PS = 1/\max_i\{t_i\} = 1/t'_i. \qquad (4)$$

On the basis of expressions (3), (4) there has been obtained the time outlay of one image processing and the through-put of the system in the dependence on the conveyor length l_k, different relation between optical and electronic devices $k_{oe}=l^o/l_k$ and on the processing time in electronic processors according to the processing time in optical ones: $a_t = t^e_{pqf}/t^o_{pjz}$.

The analysis showed that the time outlay depends proportionally inverse on the parameter value k_{oe} while having different values of l_k. On decreasing the parameter a_t the maximum time outlay T is observed at $k_{oe}=0.5$ and $l_k < 8$. The reason is that on decreasing the parameter a_t the time outlay begins to influence substantially on intermodular switching. This outlay is maximum at $k_{oe}=0.5$. Evaluation of the through-put of the system during the processing a series of images shows that on varying the value t'_i from 50 to 240ms the value PS is variable from 20 to 4 images per second.

It has been shown above that the time expenditures depends on the type of the processing device, i.e. it is much smaller in optical device. Moreover, the processing time t in the optical device does not depend on the input image, while the value of t in electronic device can be defined by a number of parameters, e.g. by the quantity of informative pixels, the image complexity. From this point of view for electronic units it is of reason to taking into account the median square deviation (MSO) value of the processing time.

Taking into account this suggestion were elaborated the new methods of design of the image processing computer systems for the next cases:

1.The value of MSO = 0 for all modules in system. This case correspond to optical system or optical electronic system, in which the processing time in the electronic modules do not depend from the input image.

2.The value of MSO $\neq 0$ for all electronic modules or part of them.

The first case correspond to the systems design in stationary mode and the next case - to the systems design in dynamic mode of image processing.

5. COMPUTER SYSTEMS DESIGN IN THE STATIONARY MODE OF IMAGE PROCESSING

Let us suppose that a large enough sequence of images will be processed, therefore the time of the first step of image loading can be neglected.

The suggested method consists of the following:

1. The image processing function W is represented as a set of operations $\{W_i\}$ in such a way that its further dividing is impossible or inexpedient. For example, such a function W_i can be indivisible, if its dividing does not lead to following decreasing of the processing time.

2. It is defined a set of operations $\{W_i\}_o$, which can be implemented with the help of optical means providing the maximum level of spatial concurrency, speed and compactness. Other functions are implemented with the help of electronic processors.

3. A corresponding switching network is forming. It consists of a buffer devices of optical, electronic types.

4. The implementation time t_{pi} of different operations W_i and the switching time t_{ki} are defined. Basing on the sets of the parameters $\{t_{pi}\}$, $\{t_{ki}\}$ a processing operation W'_i that requires the maximum time outlay is determined out:

$$t'_i = \max_i\{t_i\} = \max_i(t_{pi}+t'_{ki}).$$

5. In the case when the operation W'_i is implemented by electronic means the possibility of t'_i decreasing up to the level of other values from the set $\{t_i\}$ is examined. For this reason the possibility of introduction of spatial concurrency using a set of processor elements is being analyzed. The image is dividing into separate fragments and for processing of each of them one or several processor elements are being used.

If, however, the function W'_i is being implemented by optical means and there does not exist a possibility to decrease the parameter t'_i then it is expedient to make the justification of the parameters t_i of electronic modules till the level of t'_i by means of decreasing the number of the used processor elements or by their changing by some less rapid and therefore more inexpensive ones.

As a result we shall obtain a balanced set of the parameters $\{t_i\}$, which provides the maximum through-put of the computer system.

The described method of computer systems design assumes the absence of restrictions on the resources of processor elements in electronic modules. In the case when such limitations take place the design method will be a.f.

Let r_o from n functions, which are to be implemented, can be implemented optically and $r_e=n-r_o$ functions are to be implemented by electronic means. Besides, let the resources of electronic processors are M_r and on each step of electronic processing there M_{ri} processors are used.

Having the restrictions on processor elements (PE) resources, the following variants are possible: $r_e>M_r$; $r_e=M_r$; $r_e<M_r$. In the first case, i.e. when $r_e>M_r$ several processing operations are uniting and their implementation on one processing device are possible. If a group of operations, which are to be realized, are situated consecutively, then such way of processing is not problematical. If, however such operations are not situated consecutively then the necessity of cycle organization in the conveyor is appears. This leads to increasing of switching time outlay.

In the second and the third cases one or several processing devices can be used for each operation.

6. COMPUTER SYSTEMS DESIGN IN THE DYNAMIC MODE OF IMAGE PROCESSING

The dynamic mode of image processing is different because the time of functioning of conveyor's electronic processors is not stationary for different images and can be presented as random variable. That's why the method of system design in the dynamic mode will differ from the method of system projecting in the stationary mode. Such differences are characterized by the influence of indicated random variable on the throughput capability of the system.

In this section are presented the results of investigations of the influence of the median square deviation (MSD), the influence of time of the processing in the modules on the throughput capacity of the system under the different electronic and optical modules quantity. According to the results of investigations the recommendations of increasing the system's throughput capacity are formulated. On the bases to these recommendations, the system design method of image processing in the dynamic mode is elaborated.

6.1. The investigation of computer systems characteristics

The examined in the section 4 structural model of computer system contains the set of processing devices (modules). Each module can be presented as a separate service unit device, because the results of image processing (or processed image) can be transmitted to the next module only after the ending of calculations in the previous.

As it was presented, the time of functioning of full or partial processing in electronic devices can be examined as random variable. That's why the suggested model of the processing system can be presented as the system of mass service (SMS), where the service devices (SD) are connected consequently.

Let's T_i – is the service time of i-th service device. As the distribution of random variable T_i can be characterized by the Gauss law, it makes difficult the receiving of analytical solutions. On the other hand, because of limited capacities of buffer devices, the blocations are possible, that makes unreal the independent analysis of the each of the modules.

The mentioned conditions show that SMS is worth analyzing by the method of imitating modulation. In the process of the system's simulation it was supposed that the distribution of information processing time in the electronic modules corresponds to the normal law.

The following parameters were changing in the process of simulation:
n_e, n_o – the number of electronic, optical processing modules;
SO_e – median square deviation of processing time in the electronic modules;
m_e, m_o – median processing times in the electronic, optical modules respectively;
the coefficient of variation w, $w = SO_e/m_e$.

The measured characteristics were: throughput capability of the system $PS=1/\max\{m'_i\}$, where m'_i – is the delay time in the device; the relative throughput capability $g=(PS/PS_H-1)$, where $PS_H=1/\max\{m_i\}$ - the nominal throughput capacity.

On the first stage the influence of the relation of module number of electronic n_e and optical n_o processing in conveyor of fixed length l_k on it's throughput capability it was investigated. The values $l_k=10$, n_e was changing from 2 to 10, $n_o = l_k - n_e$; $m_e = m_o-120$; SO_e was changing from 0 to 50.
The analysis shows that if SO_e increases to 10, the value g is changing slightly. At $SO_e > 10$ the value g is lowering.

On the next stage there was investigated the influence of median processing time m_e in modules on the value g under the different parameters w, n_e. The value of parameter w was changing from 0,1 to 0,3, n_e - from 2 to 8.

The results of simulation show, that with the increasing of parameter m_e, w, n_e the throughput capability of the system is decreasing. At $w \geq 0,2$ the value g is lowering sharply.

Also the influence of median processing time m_e in the electronic modules on the value g under the different values of the parameters w, n_e were investigated. The value m_e was changing simultaneously for all electronic modules in the limits from 100 to 160. The processing time in optical modules was $m_o = 120$. The results of simulation show, that at not high value of w ($w \leq 0,1$) the parameter g slightly depends from m_e and n_e. At $w > 0,1$ and low value of n_e ($n_e = 2$) the throughput capability is minimal if $m_e = m_o$. For $m_e \neq m_o$ the throughput capability is increasing. At more high values n_e ($n_e=6$) with the increasing m_e the throughput capability is lowering.

There was investigated the influence of the value of median processing time m_o in the optical modules on the throughput capability of the conveyor for different values SO_e at $m_e=120$. The analysis shows that with the increasing m_o the g is lowering.

The results of the computer simulation show the following.

1. With the increasing of number of electronic modules in the system and of median square deviation of processing time, the throughput capability of the system is lowering.
2. At the decreasing of the median processing time in the modules and relation SO to this parameter, the throughput capability of the system is increasing.
3. The decreasing of the median processing time in the electronic modules and of the relation SO to this parameter allows also to increase throughput capability of the system.
4. The decreasing of the median processing time in the optical modules and of the processing time in the electronic modules leads to the increasing of throughput capability of the system.

6.2. The design peculiarities of the systems in the dynamical mode

In the section 5 the method of design of the optical-electronic image processing systems for stationary mode was investigated, i.e. for case when time expenses in the module's system are unchangeable while processing different images. In a number of cases the input images can differ from each other by the complexity, quantity of informational elements and other parameters, that will lead to the change of time expenses in the electronic modules. That's why the method of systems design presented in section 5, must be modified according to the results of investigations with the scope of reaching the maximal throughput capability of systems.
The analysis show that the following techniques of increasing the throughput capability are possible.

1. The maximal substitution of the electronic processing modules by optical modules, in particular of the modules with the high mead square deviation (MSD) of processing time.
2. The simultaneous reduction of the processing time in electronic and optical modules on the bases of utilization of the new algorithms and elements.
3. The reduction of the processing in the electronic devices which characterized by the high ratio of the MSD to the processing time.
4. The reduction of processing time in the optical modules.

Let's examine the design consistence of the optical electronic systems at the lack and availability of limits on the apparatus-programming resources.

The systems design without limits on calculation resources

The method of optical-electronic systems design for dynamic mode of image processing is based on materials, presented in section 5. The difference lies in the position 5. On this stage the one or several approaches, mentioned above can be used. The choice and using of one of them depends on a number of factors and leads to increasing apparatus - programming expenses.

So, the using of the first approach can be limited by the possibility of realization of the processing operation by optical means. The most effective is the second approach, because it allows to reduce the time expenses on the image processing, to increase the throughput capability. But the time expense reduction in optical and electronic modules simultaneously doesn't exist always. In this case the third or fourth approach are used.

The systems design under the limits of calculation resources

While design such systems the following approaches are possible. As in previous section, the first way can be used, i.e. the change of electronic modules on the optical, took place. Besides, it is possible the increasing of the processing time in electronic modules with low value of w under the corresponding reduction of the processing time in the modules with high value of w. Such approach is connected with redistribution of calculation resources.
If electronic modules are characterizing by not high MSD of the processing time, the increasing of time expenses in the optical modules is possible.

7. AN EXAMPLE OF IMAGE PROCESSING COMPUTER SYSTEM DESIGN

Let us investigate the efficiency of the proposed theory on the example of the design of a computer system that implements the algorithm of objects recognition based on the invariant moment features.

The algorithm consists of the following.
1.The analyzing image is being formed.
2.The noises in the image are being removed using the median filtering method.
3.The identification object is being extracted by means of threshold limitation.
4.The moment features set is calculating.
5.The object is being identified by the comparison of the calculated features with standard ones.

In this way the image processing function W can be represented as a totality of five operations W_i, i=1...5. Each of them implements a separate step of the algorithm: $P(x,y) \rightarrow W_1 \rightarrow W_2 \rightarrow W_3 \rightarrow \quad \rightarrow W_4 \rightarrow W_5 \rightarrow RO$. The operations W_2, W_4 can be implemented by optical means, W_5- only by electronic means and W_1- by optical or electronic means.

The computer system structure can be represented as it is shown in Fig.3. In this system the operation W_1 is implemented with the help of image forming unit (IFU), the operations W_2, W_3, W_4 - with the help of optical processors P_1-P_3, and operation W_5 - by an electronic processor P_4. The type of the buffer storage B_1 will depend on the type of IFU. The buffer storages B_2, B_3 are optical, but B_4 is optical-electronic.

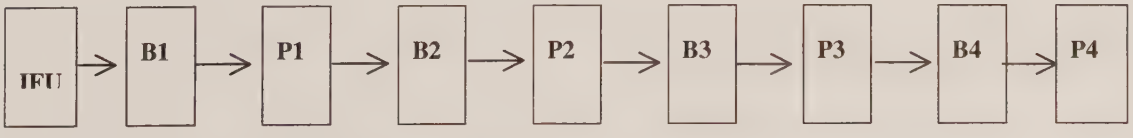


Fig.3. The Structure of an optical-electronic computer system

The implementation of this algorithm in an electronic computer system, consisting of 33 processors, linked by the hypercube topology has been investigated. The operations W_2, W_3 in such system are implemented with the help of a network of 28 processors.

The time outlay for operations implementation in the system are presented in Table 3. The analysis of data shows that the maximum time outlay in optical-electronic system is connected with the implementation of the operation W_5: $t' = t_8 + t_9 = 90ms$. The through-put of the system in this case is $P' = 1/t' = 11$ frames/sec. The maximum time outlay in a hyper cubic system is defined by the implementation of the operation W_3 and is equal to $t'' = t_2 + t_5 = 3.15sec$. That defines the through-put of the system $P'' = 1/t'' = 0.3$ frames/sec.

In such a manner the comparison of optical-electronic and hyper cubic computer systems shows that the through-put of the first of them is in 36 times higher. The total time outlay is 20 times lower.

Table 3

Time outlay in optical-electronic and hypercubic computer systems

Type of operation	Optical-electronic CS		Hypercubic CS	
	Number of processors	Processing time, ms	Number of processors	Processing time, sec
1.W^o_1		10.0		
W^e_1		40.0		1.1
2. Buffering		50.0		1.8
3.W_2	1	0.01		
4.Buffering		10.0		
5.W_3	1	0.01	28	2.07
6.Buffering		10.0		0.15
7.W_4	1	0.01	4	2.09
8.Buffering		40.0		0.2
9.W_5	1	50.0	1	0.05
The total number of processors	4		33	
Total time outlay		170.03 200.03		4.1

CONCLUSION

1. It is shown that one of the perspective ways in creating the highly productive computer means of image processing is elaboration of problem-oriented computer systems, based on combination of optical and electronic processing means. Such approach is caused that the part of operations of image processing can be realized optically with the speed, that in 3-4 levels is higher than corresponding electronic realization. At the same time there exist operations, that can be realized only with the help of electronic computer means.

2. The mathematical model of parallel image processing is suggested, that differs by the possibility of maximal parallel distribution of operations and will provide the high productiveness that realizes it's computer systems.

3. The analytical evaluation of time expenses in parallel optical-electronic image processing computer systems and the results of computer simulation, allowed to determine that while elaborating the systems of mentioned class, the length of the conveyor l_k is important, and also the relation of optical and electronic processor by the their quantity (k_{oe}) and the processing time (a_t). If $a_t > 10^4$, the increasing of throughput capability of the system is possible to achieve by choice of higher value k_{oe}. If $a_t < 10^4$, the choice of value k_{oe} it is of reason to be compared with the value of the parameter 1_k.

4. The investigation of influence of time processing characteristics in processor modules as the throughput value on the throughput capability PS system showed, that PS is increasing if the median processing time m_t in modules and the relations of median square deviation of the processing time to m_t is lowering and reduction of time processing in optical modules relatively to the time processing in electronic modules.

5. The results of the investigation of computer systems characteristics showed the following. If median square deviation of processing time SO_e in the electronic modules increases up to 10, the value of the relative throughput capability g is

changing slightly. At SO_e >10 the value g is lowering. With the increasing of the median processing time m_e in modules, the value w= SO_e/m_e and number of electronic n_e processing modules n_e in conveyor, the throughput capability of the system is decreasing. At w≥0,2 the value g is lowering sharply. At not high value of w (w≤0,1) the parameter g slightly depends from m_e and n_e. At w>0,1 and low value of n_e (n_e = 2) the throughput capability is minimal if $m_e=m_o$. For $m_e \neq m_o$ the throughput capability is increasing. At more high values n_e (n_e=6) with the increasing m_e the throughput capability is lowering.

6. The results of investigations of the influence of the median square deviation, the influence of time of the processing in the modules on the throughput capacity of the system under the different electronic and optical modules quantity permitted the elaboration of the recommendations of increasing the system's throughput capacity.

7. On the basis of the results of investigation, the methods of design the highly productive problem-oriented optical-electronic computer systems for image processing in the static and dynamic modes are elaborated.

ACKNOWLEDGMENT

The research described in this publication was made possible in part by Award No. MP2-2286 of the U.S. Civilian Research & Development Foundation for the Independent States of the Former Soviet Union (CRDF).

REFERENCES

1. Larionov A.M., Mayorov S.A., Novikov G.I. Computer complexes, systems and networks - Leningrade: Energoatomizdat, 1987.
2. Cantoni V., Levialdi S. Multiprocessor systems of image processing // Proc. IEEE. - 1988. - V.76. - No.8. - pp.118-130.
3. Artamonov G.T., Tyurin V.D. Topology of computer networks and multiprocessor systems. - Moscow: Radio i Svyaz, 1991.
4. Vasilenko G.I., Tsibulkin L.M. Holographics recognition devices. - Moscow: Radio i Svyaz, 1985.
5. Catys G.P., Perju V.L., Rotari S.L. Methods and computing means of image processing. - Kishinev: Shtiintsa 1991.
6. Radar stations of air reconnaissance/Ed. by G.S.Kondratenkov. - Moscow, 1983.
7. Perju V.L., Rotari S.L. Specialized optical-electronic computer systems of image processing // Proc. of Scientific-technical conference. - Kishinev: KPI S.Lazo, 1990. - pp.121.
8. Casasent D., Psaltis D. New optical transforms for pattern recognition//Proc. IEEE. - 1977. - V.65. - No.1. - pp.77-84.
9. Lee S., Aggarwal J. Exploitation of image parallelism via the hypercube //Proc. of the 2nd Conf. Hypercube Multiprocessors. - Knoxville, TN, Sept. 1986.-pp.426-437.
10. Casasent D., Cheatham L., Fetterly D. Optical system to computer intensity moments: design//Applied optics. -1982. - Vol.21. - No.18 - pp.3292-3298.
11. Transputers: Architecture and software // Ed. by G.Harpa - N.Y., 1991.
12. Fyodorov B.F., Elman R.I. Digital holography. - Moscow: Nauka, 1976.
13. Perju V.L. Image processing computer system construction based on optimal optical and electronic means combination. // In: Application of digital image processing. - Andrew G. Tescher, Ed. / Proc. SPIE Vol. 2564, pp.483-494 (1995).

Author Index

Alam, Mohammad S., 38, 63, 75, 214
Bauchert, Kipp, 123
Bond, A., 179
Casasent, David P., 1, 171, 265, 277, 309
Chandramouli, Lavanya, 134
Chang, Samuel, 238
Chao, Tien-Hsin, 13, 87, 109, 163, 233
Chatwin, Chris R., 51
Chen, Huaixin, 293
Chen, Xue-wen, 171
Chervonkin, Alexander P., 301
Drehman, Alvin, 115
Elkins, Les, 238
El-Saba, Aed M., 214
Ewing, Teresa, 123
Foo, Seat, 238
Fuller, J. Joseph, 238
Goh, S. F., 63
Goodhue, W., 115, 146, 288
Gregory, Don A., 223
Haider, Mohammed H., 75
Haji-Saeed, B., 115, 146, 288
Hanan, Jay C., 13, 87, 109, 233
Hassan, Hassan, 191
Horache, El-Houssine H., 63, 214
Hughlett, Casey, 87
Islam, Mohammed N., 75
Javidi, Bahram, 17
Jones, Brian K., 203
Khan, Jesmin F., 75
Khoury, Jed, 115, 146, 153, 288
Kierstead, John, 115, 146, 153, 288
Kirsch, James C., 203
Kishk, Sherif, 17
Kolluru, R., 115
Kypraios, Ioannis, 51
Lam, Raymond, 179
Lebaw, C., 179
Leon, R., 115
Li, Xiaosun, 293
Loo, Chye-Hwa, 38
Lorenzo, Joseph, 153
Mahalanobis, Abhijit, 95
Megherbi, Dalila, 146, 288
Monacos, Steve, 179
Moreels, Pierre, 233
Morell, Denise, 223
Morookian, John-Michael, 179
Nan, Jianshe, 293
Neto, Luiz Goncalves, 259
Patnaik, Rohit, 277
Perju, Veacheslav L., 309
Pizolato, Jose C., Jr., 259
Pyburn, Dana, 115, 153
Ramamoorthy, Lakshmi, 265
Regula, Srinivas, 214
Reyes, George F., 13, 109
Riasati, Vahid R., 248
Rodin, Vladislav G., 301
Sengupta, S. K., 115, 146, 153, 288
Serati, Steve, 123, 163
Shapkarina, Ekaterina A., 301
Sheikh, Tanwir, 134
Sims, S. Richard F., 95
Solyakin, Ivan V., 301
Starikov, Sergey N., 301
Talukder, Ashit, 134, 179
Tchoyprakov, Andrei, 238
Testorf, Markus, 115, 146
Thornton, Jason, 27
Vijaya Kumar, Bhagavatula V. K., 27
Vitarius, Patrick J., 223
Wang, ZhaoYang, 238
Wei, Honggang, 293
Woods, Charles L., 115, 146, 153, 288
Xia, Xiaowei, 163
Young, Rupert C. D., 51
Yuan, Chao, 1
Zhou, Hanying, 13, 87, 163
Zhou, Hui, 248
Zhou, Tong, 238

10 002736646
UNIVERSITY OF HAWAII

ISSN 0277-786X
ISBN 0-8194-5360-9